TECHNOLOGICAL APPLICATIONS OF DETERMINISTIC CHAOS

A thesis presented for the degree of
Doctor of Philosophy
in Electrical & Electronic Engineering,
in the
University of Canterbury,
Christchurch, New Zealand.

by
Alan Rodney Murch
B. E. (Hons 1)
July 1989

Abstract

Engineering and technological applications and consequences of deterministic chaos are considered. Four technological areas are investigated: electrical noise, electrical signal sources, data encryption and packet switching. In addition, deterministic chaos in both dissipative and conservative systems is reviewed, the philosophical and practical consequences that stem from deterministic chaos are considered, and the historical development that has led to the present renewal of interest in dynamical systems is presented.

The use of a hierarchy of nonlinear recursive equations to generate coloured noise and the tailoring of these equations to generate noise with a specified probability density function and power spectrum is examined. The most immediately striking aspect of the concept of a hierarchy of recursive equations is that it constitutes such an elementary means of generating a noisy process having an excess low frequency character (in particular noise having an almost 1/f power spectrum). This is claimed to constitute a significant addition to the literature on 1/f noise. An analysis of the hierarchy in terms of Lyapunov exponents and information theory is presented. It is found that the dynamics of the hierarchy exhibits nonuniformity and noise-induced predictability, thereby providing further evidence for the conjecture that nonuniformity is a necessary condition for the occurrence of noise induced predictability. It is claimed that noise-induced predictability may be of use technologically, since it might allow systems and processes which are at present unpredictable (e.g. system reliability, weather), to be made (more) predictable and useful.

The extent and under what circumstances deterministic chaos contributes to noise within sinusoidal oscillators is considered. Effects (i.e. signal-dependent delay and circuit parameter variations) which tend to be neglected in established approaches to oscillator analysis are included in the oscillator models studied. Conditions sufficient for an oscillator to exhibit deterministic chaos are found to be, first, the existence of a signal delay within the oscillator and, second, certain types of amplifier nonlinearity. It is conjectured that all oscillators may satisfy these conditions, and therefore to some degree exhibit deterministic chaos. A novel discrete time oscillator is developed. It gives insights into the way deterministic chaos arises within a closed loop. Chaotic behaviour in generalisations of an oscillator circuit named after Chua are examined. It is found that relatively minor circuit alterations can inhibit the chaotic behaviour in the circuit. It is indicated how this appears to provide insight into chaotic dynamics in general.

The extent to which the seemingly random numbers generated by a chaotic dynamical system are suitable for data encryption is examined. This leads to an examination of the consequences of number quantisation (i.e. the use of finite precision

numbers) which inevitably occurs in digital implementation of chaotic systems. Two encryption schemes are considered, termed isolated and influenced chaotic encryption. Isolated chaotic encryption is based on a conventional encryption scheme (the one time tape), while influenced chaotic encryption is a new encryption method. Sequences exhibiting maximal entropy are required for encryption. Computer simulations performed on finite length sequences (limited to allow the computer simulation to run in a reasonable time), generated by the encryption schemes, show that they exhibit maximal entropy. Although this does not confirm that sequences longer than those tested exhibit maximal entropy, it does reveal that deterministic chaos has potential for data encryption and is worthy of further study. The advantage of chaotic encryption schemes over other schemes may lie in their ease of implementation, and the difficulty of breaking the encryption from observation of the encrypted sequence.

A new packet switching flow control algorithm, termed cooperative flow control, is examined. The new algorithm is based on a modification to an algorithm which forms the basis of many flow control algorithms used in practice. Cooperative flow control represents an attempt at developing a packet entry flow control algorithm which induces resource efficient self-organising behaviour within a packet switching network, while at the same time ensuring that a specified grade of service is delivered to network users. Under certain traffic arrival patterns cooperative flow control becomes chaotic, demonstrating that the algorithm can induce self-organising behaviour within the network. The studies reported herein suggest that if such an algorithm can be perfected it may provide a considerable advance in communication network design.

Acknowledgements

I am indebted to my supervisor, Professor R. H. T. Bates, for his insight, enthusiasm and enormous tolerance. Through his efforts I have witnessed the excitement of engineering and scientific endeavour.

I am also indebted to my family, especially my parents and my brother Ross, who were always there, and provided unconditional support when I needed it.

I am privileged to have been able to work with and along side the staff and students of the Electrical Engineering Department. You have, and will always be an inspiration to me.

I am grateful to all those people who knowingly or unknowingly provided help and support at crucial moments during my studies. I could not have done it without you.

I gratefully acknowledge the generous financial support provided by Telecom Corporation of New Zealand Ltd.

Contents

Preface

This century has seen three revolutions in the science of dynamics. The first two were relativity and quantum mechanics. These changed the laws of physics under conditions of speed and size far removed from our direct experience. The third revolution is the new insights into complexity, which has left Newton's laws of motion unaffected, but has radically altered our understanding of the behaviour they describe. This third revolution has stemmed from the use of computers in the study of nonlinear physics over the past thirty years. Such studies have introduced two new theoretical constructs into the field of dynamics. The first is the soliton, and the second is deterministic chaos. This thesis is concerned with the second construct, deterministic chaos.

There has been an explosive growth in the number of researches into deterministic chaos in recent years. Although this field has fuelled a small revolution in scientific thought, little in the way of direct useful consequences and applications to the engineering and technological sciences has occurred. This situation is possibly a reflection of the general unawareness of the true nature of deterministic chaos amongst engineers and technologists. There have been a number of amusing and interesting documented cases where such an awareness might have proved useful. Perhaps the most classic example is provided by van der Pol and van der Mark (1927) when they noted that, for particular circuit values, an oscillator they were studying generated unexpectantly random, or as they describe, "irregular noise". Kennedy and Chua (1986) have recently shown that this "irregular noise" is caused by deterministic chaos. Such noise turns out in many cases to be the main rather than the subsidiary (as van der Pol and van der Mark believed) phenomenon.

Deterministic chaos manifests itself whenever certain types of nonlinearities govern the dynamics of a device, be it a system or a circuit. For certain combinations of the device parameters, the nonlinearities may impede the operation of the device by generating unanticipated phenomena. Yet the same nonlinearities are usually essential for the unique properties and the desired operation of such devices. As an example, think of the process of designing a nonlinear oscillator to generate a single specific frequency. In addition to generating the desired frequency, the nonlinearity may give rise to subharmonic oscillations or deterministic chaos. The parameters specifying the operation of the nonlinear oscillator are critical. For a certain parameter range, the oscillator may generate the desired single frequency, while for other parameter ranges it may not. Consequently, in order to design an oscillator free from "irregular noise", it is important to identify the parameter ranges for which chaos-free operation is guaranteed. The traditional tools employed in design cannot conveniently be used in these circumstances because they assume a particular type of operation (e.g. oscillation at a single frequency) and/or because of the type of non-

linearity present. Often in these cases, it is easier to identify the parameter range of acceptable operation by identifying its complement (i.e. by identifying the parameter range over which deterministic chaos arises). Moreover, the latter approach tends to educate us about the character, the features, and the properties of the deterministic chaos. This viewpoint underscores the potential engineering significance of chaos, assisted by computer calculations and simulations, in both the analysis and the design of circuits and systems.

Chaos is a word of Greek origin that denotes the primeval god from whom everything originated. The children of Chaos were Darkness, Night and Fate. Then came Heaven and Earth and the other gods of Olympus. The Greeks considered Chaos as preceding order. Today chaos takes on an almost opposite perspective. The study of deterministic chaos is the study of finding order within chaos. After an initial beginning with an air of unrespectability and as a trendy research topic, chaos has definitely come of age and is now considered by some as one of the most exciting and important research areas. The term 'chaos' was coined quite recently, by Li and York (1975), to describe apparent stochastisty in deterministic systems. Since then the literature has almost universally adopted the term. In this thesis the term 'chaos' and 'deterministic chaos' are used interchangeably.

The purpose of this thesis is to examine potential applications and consequences of deterministic chaos in technologically important areas. Four technological areas are investigated: electrical noise, electrical signal sources, data encryption and packet switching. Two applications of deterministic chaos are described: a method for generating noise with a specified probability density function and power spectrum, and encryption of data. Chapters 1,2,3 and 4 provide the theoretical basis and sets the scene for the later Chapters. Chapters 5,6,7 and 8 include new results. Each Chapter introduces additional review material where required. Conclusions and suggestions for further work on the topics discussed in Chapters 5,6,7 and 8, are collected in Chapter 9.

Chapter 1 reviews deterministic chaos in both dissipative and conservative systems, and has three goals: 1) to give an intuitive feel for how complicated behaviour can arise in deterministic systems, 2) to introduce many of the terms used in the specialist literature, 3) to provide the theoretical basis for later Chapters. Chapter 1 places more emphasis on intuition than rigour and attempts to expose the interrelationship between conservative and dissipative systems, which few other reviews have attempted. The Chapter begins with a general discussion of dynamical systems. This provides the basis for splitting the Chapter into the two sections: dissipative systems and conservative systems. The section on dissipative systems discuss the reasons why it is difficult to precisely define the concepts now known as attractors and deterministic chaos. The section on conservative systems begins with an explanation of the fundamental difference between conservative and dissipative systems and ends with an intuitive description of KAM (Kolmogorov, Arnold, Moser) theory. Many of the general mathematical terms and concepts required to provide the proper setting for explaining deterministic chaos are introduced in Chapter 2. This rids Chapter 1 of the disruption caused by the need to continually define new terms and concepts. It also makes for less irritating reading for those who are already familiar with these ideas. Chapter 3 outlines some of the philosophical and practical consequences that have stemmed from deterministic chaos. It is explained how improved understanding of nonlinear phenomenon in general has lead to some surprising consequences, partic-

ularly those suggested by Prigogine. The important practical consequences of chaos are discussed with reference to experimental observations reported in the literature. Chapter 4 is devoted to the historical development of those dynamical systems which have inspired the present renewal of interest in deterministic chaos. The more important events that occurred in the seventeenth, eighteenth and nineteenth centuries are briefly described. The bulk of the discussion is of developments in the 1960s and later.

Chapter 5 introduces a computational framework seemingly capable of generating sequences of numbers exhibiting arbitrary probability density functions and power spectra. This might permit a simple and versatile coloured noise generator to be implemented with some ease. The most immediately striking aspect is that it constitutes such an elementary means of generating a noisy process having an excess low frequency character. It might be conjectured that this constitutes a significant addition to the literature on 1/f noise, because its very simplicity may suggest a hitherto overlooked physical explanation for the wide occurrence of such noise. Chapter 6 assesses to what extent the noise in the output of a high quality sinusoidal oscillator is attributable to deterministic chaos. The noise performance of a typical real-world sinusoidal oscillator seems to be somewhat worse than established approaches to noise analysis of theoretical models of such oscillators would suggest (Robins 1984, page 63). Effects which tend to be neglected in such analysis are included in the oscillator models developed in Chapter 6. Chapter 7 assesses to what extent the seemingly random numbers generated by chaotic dynamical systems are suitable for data encryption. Chaotic dynamical systems have properties which appear highly desirable for secure data transmission systems. There does not seem to be any specific mention of chaotic dynamical systems for data encryption in the literature. Chapter 8 analyses a packet switching flow control algorithm which can become chaotic under certain input traffic patterns. The development of a packet entry flow control algorithm that induces resource efficient self-organising behaviour, while at the same time delivering the specified service to users, may provide a considerable advance in communication network design. The development of such an algorithm is attempted in Chapter 8. This algorithm is based on a modification to an algorithm which forms the basis of many flow control algorithms used in practice (Gerla and Kleinrock 1980).

A decimal system of section numbering is adopted in this thesis, and sections are referred to in the text by the symbol § followed by the section number. Thus §2.3.1 refers to the first subsection of the third section of Chapter 2. Equations are numbered consecutively within each Chapter and are referred to by enclosing the equation numbers in parentheses. Thus (4.2) refers to equation two in Chapter 4. Figures are numbered in the same manner as equations, and are referred to by preceding the figure number with 'figure'. Thus figure 2.4 refers to figure four in Chapter 2. When a new term is defined or introduced for the first time, it is *emphasized* in italicized type.

During the course of the work reported in this thesis the following papers and presentations have been prepared.

A. R. Murch, W. K. Kennedy and R. Davidson, Traffic and Performance Measurements on PACNET, Proceedings of The National Electronics Conference, Auckland, Vol. 24, pp. 107-110, 1-3 September, 1987.

A. R. Murch and R. H. T. Bates, Non-Random Noise Mechanisms, Proceedings of The National Electronics Conference, Auckland, Vol. 24, pp. 137-140, 1-3 September, 1987.

R. H. T. Bates and A. R. Murch, Deterministic-Chaotic Variably Coloured Noise, Electronic Letters, Vol. 23, No. 19, pp. 995-996, 10 September, 1987.

V. A. Smith, A. R. Murch and A. Dingle, Non-Random Noise Mechanisms, Proceedings of The National Electronics Conference, Christchurch, Vol. 25, pp. 188-193, 31 August-2 September, 1988.

A. R. Murch and R. H. T. Bates, Colored Noise Generation Through Deterministic Chaos, Accepted for publication in IEEE Transactions on Circuits and Systems.

Chapter 1

Introduction to Dynamical Systems

> Newton's fundamental discovery, the one which he considered necessary to keep secret and published only in the form of an anagram, consists of the following: *Data aequatione quotcunque fluentes quantitae involvente fluxions invenire et vice versa.* In contemporary mathematical language this means: "It is useful to solve differential equations" (Arnold 1983, page iii).

> ...the crude speculation that all dynamical systems are periodic or nearly so presents itself irresistibly to the human mind (Birkhoff 1941).

Scientific method is essentially a process for finding patterns within systems (natural or man-made). A *system* is something having parts which is perceived as a single entity, and a dynamical system is one which changes or evolves with time (Hirsch 1984, page 3). Patterns are studied by developing models for these systems, and such models are described by mathematical equations called *system equations*. The system equations for dynamical systems are usually *difference equations* and/or *differential equations* (Hirsch 1984, page 6). The types of systems reviewed in this Chapter can mainly be characterised by ordinary differential equations (ODE) and difference equations arising from ODEs.

A *dissipative dynamical system* is one that squanders energy. It releases energy into its environment which it cannot reuse. In general, such a system eventually loses all its energy and finally comes to rest. If enough external energy is supplied, a dissipative dynamical system may not necessarily come to rest. Its equilibrium behaviour may be periodic motion or something more complicated (assuming the system remains bounded). Once equilibrium is reached, an external perturbation may alter the behaviour but this can only be transient and the system must return to its equilibrium state. Dissipative systems are therefore robust in that they tend (or are attracted) to a particular behaviour. This is how dissipative systems fundamentally differ from *conservative dynamical systems* which (by definition) conserve energy (conservative systems are discussed in §1.2). Dissipative systems are said to possess attractors (refer to §1.1.3) whereas conservative systems do not. These ideas are made clearer by the following example. A frictionless pendulum is a conservative system because, once set in motion, it swings forever, repeating the same pattern. If the

pendulum is perturbed (bumped), the motion adopts a new pattern, which it retains ever afterwards, provided it suffers no further disturbance. Suppose the system is made dissipative by introducing friction to damp the motion of the pendulum. It necessarily dissipates energy in the form of heat. Eventually the pendulum comes to rest. If instead the oscillations are sustained by driving the pendulum with a periodic force, which supplies external energy, it adopts the period of motion imposed by the external forcing supply. Under these conditions, any disturbance to the motion is only transient, being dissipated away by friction. The driving force soon reasserts its characteristic pattern of motion. The dissipation flushes out disturbances and endows such a system with robustness and stability (Davies 1987, page 42).

The study of dynamical systems had its origins in the study of the solar system (Hirsch 1984). While the solar system is not actually a conservative system, since it dissipates energy (tidal forces, etc., actually cause it to loss energy, but these losses can be considered negligible even over quite long periods), the solar system is effectively an example of a conservative dynamical system. The study of this particular system preoccupied many mathematicians and scientists for hundreds of years. This resulted in the nature of conservative systems becoming reasonably well understood during the 1800s by mathematicians such as Laplace and Poincaré (Hirsch 1984). In the late 1800s it was known that certain conservative systems could behave in very complicated ways. Only in this century has very complicated behaviour in dissipative systems become comprehensively studied. The mathematical methods developed for analysing conservative and dissipative systems are similar, but have important differences. For this reason the study of conservative systems has developed separately from that of dissipative systems.

Due to the differences between dissipative and conservative systems, this Chapter is split into two sections entitled: dissipative dynamical systems and conservative dynamical systems. Most of the literature on deterministic chaos is due to professional mathematicians and theoretical physicists. Much of this literature is effectively incomprehensible to workers in other disciplines due to its specialist nature. Most reviews (although not all *cf.* Guckenheimer *et al.* 1977; Kloeden and Mees 1985; Hirsch 1984) on deterministic chaos seem either to be non-technical (*cf.* Hofstadter 1981; Ford 1983; Crutchfield *et al.* 1986; Gleick 1987) or to provide too narrow a perspective of the topic to be widely useful (*cf.* Collet and Eckmann 1980; Tomita 1982; Eckmann and Ruelle 1985; Chernikov *et al.* 1988). This Chapter is intended to interpret parts of the specialist literature for engineers and various other applied scientists. Furthermore, this Chapter has three particular purposes: 1) to give an intuitive feel for how complicated behaviour can arise in deterministic systems, 2) to introduce many of the terms used in the specialist literature, 3) to provide the theoretical basis for later Chapters. Many of the general mathematical terms and concepts required to provide the proper setting to explain deterministic chaos are introduced in Chapter 2. This rids Chapter 1 of the disruption caused by the need to continually define new terms and concepts. It also makes for less irritating reading for those who are already familiar with these ideas. Whenever a new term or concept is introduced the reader is referred to the appropriate section (say the m^{th}) in Chapter 2 by enclosing 'refer to §2.m' in parentheses.

1.1 Dissipative Dynamical Systems

The theory of ordinary differential equations is one of the basic tools of mathematical science. The theory allows the study of all kinds of evolutionary systems with the properties of *determinacy*, *finite-dimensionality*, and *differentiability* (Arnold 1973, page 73).

A system is said to be deterministic if its entire past and future are uniquely determined by its present state (*cf.* Pippard 1985). The set of all possible states of a system is called its *state space*. Thus, for example, classical mechanics considers systems whose past and future are uniquely determined by the initial positions and velocities of all particles within the system. The state space is just the set of instantaneous positions and velocities of the particles. Heat propagation is a semi-deterministic system in the sense that its present uniquely determines its future, but not its past. Quantum mechanics is non-deterministic in the sense that it only predicts the probabilities of particles following particular trajectories.

A system is said to be finite-dimensional if its state space possesses a finite number of dimensions (*cf.* Pippard 1985). Systems which are not finite-dimensional include propagation of waves in optics and acoustics, motion of fluids, and systems exhibiting time delays. The former systems are described by partial differential equations (PDE) while the latter system is described by ODEs which include terms specifying the present and (at particular instants in the) past states of the system. Such ODEs are known as ODEs incorporating time delays, or time delay ODEs.

A system is said to be differentiable if its time-varying state can be described by a differentiable function (i.e. the motion is smooth) (*cf.* Pippard 1985). A considerable theory for systems which are not differentiable has been developed. Examples of such theories include the theory of shock waves, elasticity, optics, acoustics and catastrophe theory (Gilmore 1981).

The *order* of an ODE is the order of the highest derivative that occurs explicitly in the equation (Braun 1983, page 1). The order of an ODE determines the dimensionality of its state space, which always equals the order of the ODE. An n^{th}-order ODE can always be transformed into a set or system of n one-dimensional equations, i.e. (Braun 1983, page 263)

$$\frac{d\mathbf{x}}{dt} \stackrel{def}{=} \dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) \tag{1.1}$$

where $\mathbf{x}$ and $\mathbf{f}$ are vectors, with $\mathbf{x}$ depending upon a real scalar variable t (i.e. $\mathbf{x} = \mathbf{x}(t)$). If the vector $\mathbf{f}$ is a function of $\mathbf{x}$ only (i.e. not explicitly dependent on t) then (1.1) is called an *autonomous system*, otherwise (1.1) is called a *non-autonomous system* (i.e. $\mathbf{f}$ depends explicitly on $\mathbf{x}$ and t) (Chua 1987, page 982). The set of all n-tuples of real numbers is called real n-dimensional *Cartesian space* and is denoted $\mathbf{R}^n$ (Chinn and Steenrod 1966, page 9). An element of $\mathbf{R}^n$ is a point $\mathbf{x} = (x_1, \cdots, x_n)$, where the number x_i is the i^{th} *coordinate* of the point $\mathbf{x}$ (refer to §2.1). Each element of $\mathbf{R}^n$ specifies a state of the system and $\mathbf{R}^n$ is the space containing all possible states. Such a space is called the *state space* or *phase space* of the system (Hirsch and Smale 1974, page 22). The function $\mathbf{x}(t)$ in (1.1) specifies an unknown curve $\mathbf{x}(t) = (x_1(t), \cdots, x_n(t))$ in the state space $\mathbf{R}^n$. That is, $\mathbf{x}(t)$ is a mapping from the real numbers $\mathbf{R}$ into $\mathbf{R}^n$ (i.e. $\mathbf{x} : \mathbf{R} \rightarrow \mathbf{R}^n$, refer to §2.3). Initial conditions are of the form $\mathbf{x}(t_0) = \mathbf{u}$ where $\mathbf{u} = (u_1, \cdots, u_n)$ is a point in $\mathbf{R}^n$. Geometrically this means

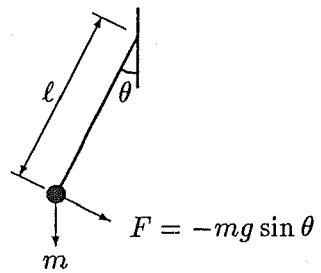


Figure 1.1: Representation of a pendulum. The pendulum is of length ℓ, with a bob mass of m.

that when $t = t_0$, the curve is required to pass through the given point $\mathbf{u}$. Solving the differential equation (1.1), with initial conditions $\mathbf{u}$, means finding a solution curve $\mathbf{x}(t)$ that satisfies (1.1) and passes through the point $\mathbf{u}$ when $t = t_0$ (Hirsch and Smale 1974, page 4). To indicate the dependence of the solution curve on t and $\mathbf{u}$, it is denoted by $\mathbf{x}(t, \mathbf{u})$.

Consider (1.1) as a dynamical system. This means that the real scalar variable t is interpreted as time (Hirsch and Smale 1974, page 5) and the solution curve $\mathbf{x}(t, \mathbf{u})$ could be thought of, for example, as the path of a particle moving in $\mathbf{R}^n$. The function $\mathbf{f}$ in (1.1) is often called the *velocity function*, since it specifies the rate of change of $\mathbf{x}$. The term velocity is carried over from classical mechanics (Hirsch and Smale 1974, page 287). The solution curves of dynamical systems are called *trajectories* or *orbits* (Hirsch and Smale 1974, page 5). As time proceeds every point in $\mathbf{R}^n$ moves simultaneously along the trajectory passing through it. In this way the collection of maps $\mathbf{x} : \mathbf{R}^n \to \mathbf{R}^n$, $t \in \mathbf{R}$, becomes a one parameter family of transformations. This set of all trajectories $\mathbf{x}(t, \mathbf{u})$ is called the *flow*, or the *dynamics*, or the *dynamical system*, and is denoted $\phi_t(\mathbf{u})$ (Hirsch and Smale 1974, page 6). The flow could be imagined as the trajectories traced out by particles placed at each point in $\mathbf{R}^n$ all moving simultaneously with t (e.g. the trajectories traced out by dust particles under a steady wind). Due to constraints on the system (e.g. limited energy, physical barrier, etc.) the flow is usually only defined within a subset of state space. This subset is termed the *allowed region* of state space. The flow $\phi_t(\mathbf{u})$ obeys the following rules (called *group properties*) (Hirsch and Smale 1974, page 174)

$$\begin{aligned} \phi_0(\mathbf{u}) &= \text{identity} \\ \phi_s(\mathbf{u}) \circ \phi_t(\mathbf{u}) &= \phi_{s+t}(\mathbf{u}) \end{aligned} \tag{1.2}$$

where the symbol $\circ$ indicates composite function (i.e. $\phi_s(\mathbf{u}) \circ \phi_t(\mathbf{u}) = \phi_s(\phi_t(\mathbf{u}))$, refer to §2.3).

Consider as an example a frictionless pendulum of length ℓ and of bob mass m; refer to figure 1.1. If the pendulum is swung to an angle θ (where θ is the angular displacement of the pendulum from its rest position), the component F of force tangent to the arc traced out by the bob (i.e. the component of force trying to restore the pendulum to its vertical position) is given by

$$F = -mg \sin\theta \tag{1.3}$$

where g is the universal gravitation constant. By convention the restoring force is negative. The angular acceleration $a = \ell\ddot{\theta}$ of the bob induced by the restoring force is

$$\begin{aligned} a &= \ell\ddot{\theta} = \frac{F}{m} \\ \Rightarrow \ddot{\theta} &= -\frac{g}{\ell}\sin\theta \end{aligned} \tag{1.4}$$

This 2nd order ODE can be written as a system of two first order ODEs, by setting $x_1 = \theta$ and $x_2 = \dot{x_1} = \dot{\theta}$, and can also be normalized by setting $g/\ell = 1$, thus

$$\begin{aligned} \dot{x_1} &= x_2 \\ \dot{x_2} &= -\sin x_1 \end{aligned} \tag{1.5}$$

A geometrical picture of the flow can be set up in state space. Figure 1.2 plots some of the trajectories in state space for the system represented by (1.5). The arrows indicate the direction of the trajectories as time increases. The flow is the collection of all trajectories. Geometrically the flow can be visualised by a few trajectories. Such a diagram is called a *state portrait* or *phase portrait* (Braun 1983, page 416). Figure 1.2 is the state portrait for the above example.

A dynamical system always has a certain number of *degrees of freedom*, the number of which can be equal to or less than the dimension of the dynamical system's state space (Goodman and Warner 1964, page 142). If the state space is n-dimensional and each coordinate of state space can change independently then the system has n degrees of freedom (Goodman and Warner 1964, page 339). If there are constraints on the system that prevent each coordinate changing independently then the number of degrees of freedom is less than the dimension of state space. Again consider, as an example, the frictionless pendulum of length ℓ and of bob mass m. The total energy of the system remains constant, since the frictionless pendulum is a conservative system. As the pendulum swings, kinetic energy is transformed to potential energy and vice verse. Let the total energy of the system be represented by H. The kinetic energy T is

$$T = \frac{1}{2}m(\ell x_2)^2 \tag{1.6}$$

where $\ell x_2 = \ell\dot{\theta}$ is the angular velocity of the bob. The potential energy V is

$$V = -mg\ell\cos x_1 \tag{1.7}$$

The potential energy V is a maximum when the bob is standing vertically upwards and a minimum when the bob is sitting vertically downwards. Any constant added to V does not alter the equations of motion but only changes the potential energy reference. The total energy is

$$\begin{aligned} H &= T + V \\ &= \frac{1}{2}m\ell^2 x_2^2 - mg\ell\cos x_1 \end{aligned} \tag{1.8}$$

where x_1 is the angular displacement of the pendulum from its rest position (i.e. the vertically downward position) and g is the universal gravitation constant (note that,

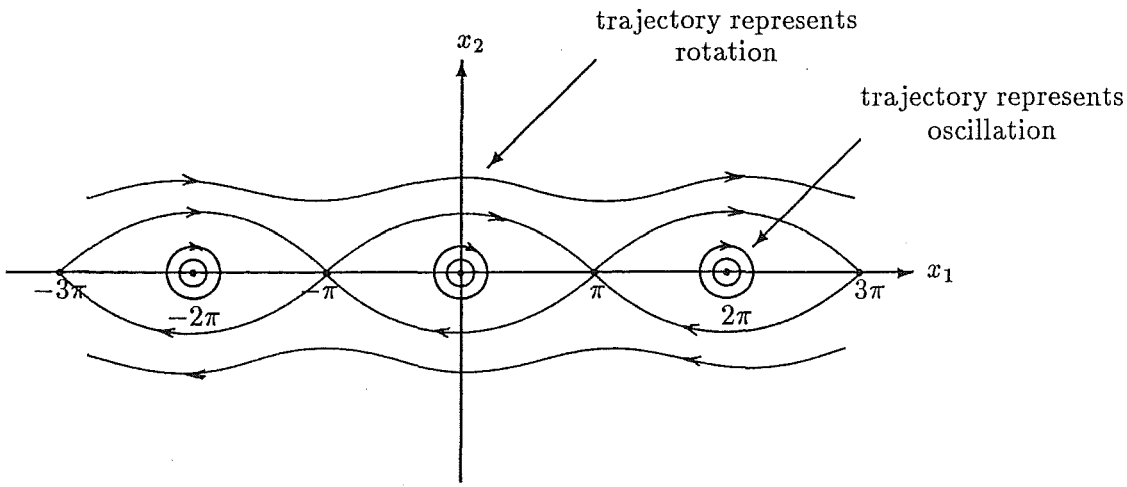


Figure 1.2: State portrait of a frictionless pendulum characterised by the dynamical system represented by (1.5).

H is called the Hamiltonian; refer to §1.2.1). The state space is two-dimensional and possesses coordinates x_1 and x_2. Since the energy of the pendulum is constrained to be constant, when one coordinate of state space is specified then the other is as well. Thus the system has one degree of freedom. In general, the number of degrees of freedom is equal to the dimension of state space reduced by the number of constraint relations.

The long term or *asymptotic behaviour* (i.e. the behaviour approached asymptotically as $t \to \infty$) of many dynamical systems is often contained within a region of state space that possesses a small number of dimensions (i.e. fixed points, periodic orbits, etc.; refer to §1.1.2) (Guckenheimer and Holmes 1983, page 33). That is, many of the coordinates of state space are not required in a description of the asymptotic behaviour. This also appears to be true for dynamical systems which exhibit seemingly chaotic behaviour. If this phenomenon is true in general, it is perhaps fortunate, since it may be possible to model complicated systems with (potentially) many degrees of freedom by simpler systems having fewer degrees of freedom.

An important notion in the study of dynamical systems is the stability or persistence of a system under small changes or perturbations of the system equations. This concept is called *structural stability* (Hirsch and Smale 1974, page 304). If the system equations of a dynamical system are perturbed slightly in any way, then the perturbed dynamical system is said to lie nearby the unperturbed dynamical system (this is made precise in §2.3). A dynamical system is structurally stable if every nearby dynamical system is completely equivalent in qualitative terms, or more precisely, is *topologically equivalent* (refer to §2.3). Since assumptions, approximations and experimental error are always present in any physical model, the system equations describing the model, while providing a completely accurate solution to the physical model, are nevertheless only an approximation to reality since the model itself suffers this flaw. If the dynamical system is not structurally stable, the small errors and approximations made in the model have a chance of dramatically changing the structure of the real solution to the system.

1.1.1 The Linear System

A linear system of differential equations is a special case of (1.1), where the function $\mathbf{f}$ is a polynomial of first degree in the variables $x_1, \cdots, x_n$ (Coddington and Levinson 1955, page 108). Thus a linear system can be written in matrix form as

$$\dot{\mathbf{x}} = A\mathbf{x} \tag{1.9}$$

where A is a $n \times n$ matrix with constant coefficients. The importance of linear systems is that their behaviour is nowadays completely understood, since they can be fully described in terms of the highly developed methods of *linear algebra* (Hirsch and Smale 1974). The solution curve to (1.9) is given by

$$\mathbf{x}(t, \mathbf{u}) = e^{tA}\mathbf{u} \tag{1.10}$$

where e^{tA} is the $n \times n$ matrix obtained by exponentiating the matrix A (Guckenheimer and Holmes 1983, page 8). Linear systems can be classified into distinct types depending on the eigenvalues of A (Percival and Richards 1982, page 33). For a 2×2 matrix there are 3 distinct types:

Type 1: the two eigenvalues of A are real and distinct.

Type 2: the two eigenvalues of A are complex conjugates of each other.

Type 3: the two eigenvalues of A are real and equal.

The matrix A can be simplified by invoking a non-singular matrix M to introduce a linear change of coordinates (i.e. $\mathbf{x} = M^{-1}\mathbf{y}$, $\mathbf{y} = M\mathbf{x}$ where $\mathbf{y}$ is the new coordinate), so that

$$\dot{\mathbf{y}} = MAM^{-1}\mathbf{y} = B\mathbf{y} \tag{1.11}$$

where $\mathbf{y}$ is the new coordinate (Hirsch and Smale 1974, Chapter 3). Matrices A and B are topologically equivalent which means the dynamics of $\dot{\mathbf{y}} = B\mathbf{y}$ is completely equivalent in qualitative terms to the dynamics of $\dot{\mathbf{x}} = A\mathbf{x}$ (refer to §2.3). It is possible to choose a matrix M, depending on the elements of A, so that B is of a simple standard form. For each distinct type of linear system there corresponds a standard matrix form B (Percival and Richards 1982, Chapter 3). The three standard forms for 2×2 matices are

$$\begin{pmatrix} \hat{\lambda}_1 & 0 \\ 0 & \hat{\lambda}_2 \end{pmatrix}, \begin{pmatrix} \alpha + j\beta & 0 \\ 0 & \alpha - j\beta \end{pmatrix}, \begin{pmatrix} \hat{\lambda} & 0 \\ c & \hat{\lambda} \end{pmatrix} \tag{1.12}$$

where $\hat{\lambda}_1 \neq \hat{\lambda}_2$ and $c \in \mathbf{R}$. Illustrative state portraits associated with each of the three standard matrix forms (1.12) are shown in figure 1.3.

The standard forms for general $n \times n$ matrices are called *Jordan forms* (Hirsch and Smale 1974, page 126). Let the eigenvalues of a general $n \times n$ matrix be denoted $\hat{\lambda}_j, j = 1, \cdots, n$ and the eigenvectors of A be denoted $v_j, j = 1, \cdots, n$. The matrix M is constructed from the eigenvectors of A, which forms the columns of M (Guckenheimer and Holmes 1983, page 9). If A has distinct eigenvalues (i.e. is of type 1 or 2) the matrix $B = MAM^{-1}$ is diagonal and the diagonal elements of B are the

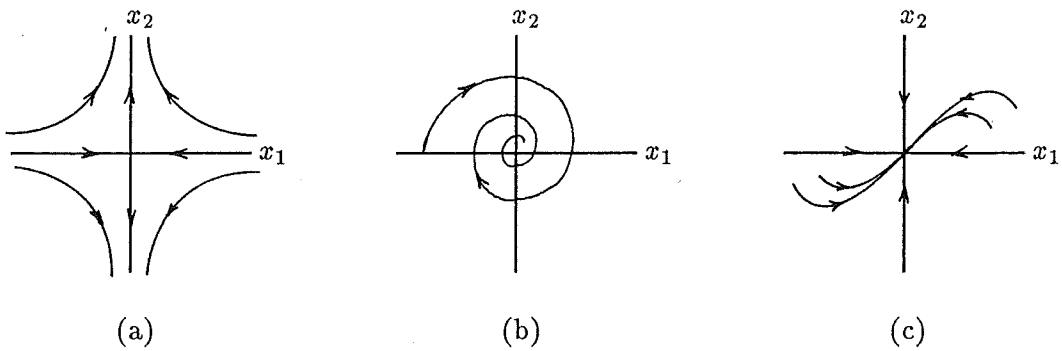


Figure 1.3: Illustrative examples of state portraits associated with each of the three standard matrix forms given by (1.12). **(a)** Type 1. **(b)** Type 2. **(c)** Type 3.

eigenvalues of the A matrix (Hirsch and Smale 1974, page 46). The solution curve of the transformed linear system $\dot{\mathbf{y}} = B\mathbf{y}$ is

$$\begin{aligned} \mathbf{y}(t, \mathbf{y}(t_0)) &= e^{tB}\mathbf{y}(t_0) = \begin{pmatrix} e^{\hat{\lambda}_1 t} & & \\ & \ddots & \\ & & e^{\hat{\lambda}_n t} \end{pmatrix} \mathbf{y}(t_0) \\ \Rightarrow y_i(t, \mathbf{y}(t_0)) &= e^{\hat{\lambda}_i t} y_i(t_0) \end{aligned} \tag{1.13}$$

where $y_i(t_0)$ is the i^{th} component of $\mathbf{y}(t_0) = M\mathbf{u}$. The solution curve $\mathbf{x}(t, \mathbf{u})$ is

$$\begin{aligned} \mathbf{x}(t, \mathbf{u}) &= M^{-1}\mathbf{y}(t, \mathbf{y}(t_0)) = M^{-1}e^{tB}\mathbf{y}(t_0) = M^{-1} \begin{pmatrix} e^{\hat{\lambda}_1 t} & & \\ & \ddots & \\ & & e^{\hat{\lambda}_n t} \end{pmatrix} \mathbf{y}(t_0) \\ &= \sum_{j=1}^{n} y_j(t_0)\mathbf{x}_j(t) \quad \text{where} \quad \mathbf{x}_j(t) = e^{\hat{\lambda}_j t} v_j \end{aligned} \tag{1.14}$$

The coordinate transformation MAM^{-1} rotates each eigenvector of the A matrix until it coincides with a different coordinate axis of state space. This process is called *uncoupling* (Hirsch and Smale 1974, page 67). The state portrait of an uncoupled system can easily be drawn. For example, consider a linear system where the A matrix has real distinct eigenvalues

$$\dot{\mathbf{x}} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mathbf{x} \tag{1.15}$$

After a linear change of coordinates $\mathbf{x} = M^{-1}\mathbf{y}$, it is possible to diagonalise the A matrix

$$\dot{\mathbf{y}} = \begin{pmatrix} \hat{\lambda}_1 & 0 \\ 0 & \hat{\lambda}_2 \end{pmatrix} \mathbf{y} \tag{1.16}$$

and the solution is

$$\mathbf{y}(t, \mathbf{y}(t_0)) = \begin{pmatrix} e^{\hat{\lambda}_1 t} & 0 \\ 0 & e^{\hat{\lambda}_2 t} \end{pmatrix} \mathbf{y}(t_0) \tag{1.17}$$

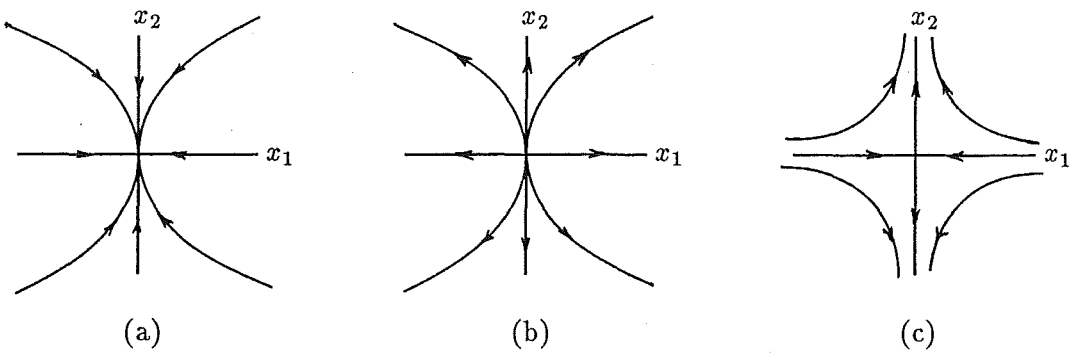


Figure 1.4: Illustrative state portraits for (1.17), **(a)** $\hat{\lambda}_1 < \hat{\lambda}_2 < 0$. **(b)** $0 < \hat{\lambda}_1 < \hat{\lambda}_2$. **(c)** $\hat{\lambda}_1 < 0 < \hat{\lambda}_2$.

which is a system of independent equations

$$\begin{aligned} y_1\left(t, \mathbf{y}(t_0)\right) &= e^{\hat{\lambda}_1 t} y_1(t_0) \\ y_2\left(t, \mathbf{y}(t_0)\right) &= e^{\hat{\lambda}_2 t} y_2(t_0) \end{aligned} \tag{1.18}$$

the state portrait of which is shown in figure 1.4(a)-(c), for $\hat{\lambda}_1 < \hat{\lambda}_2 < 0$, $0 < \hat{\lambda}_1 < \hat{\lambda}_2$ and $\hat{\lambda}_1 < 0 < \hat{\lambda}_2$ respectively.

Note from the above state portraits that the trajectory of every point lying on the subspace spanned by an eigenvector (refer to §2.2) remains in that subspace for all forward and reverse time (i.e. for $t \to \infty$ and $t \to -\infty$). Such a subspace is called invariant under the flow $\phi_t(\mathbf{u}) = e^{tA}\mathbf{u}$ (Guckenheimer and Holmes 1983, page 10). In general, if $\{v_j : j = 1, \cdots, k\}$ for $k \leq n$, is any set of real eigenvectors of A, a point on the subspace spanned by $\{v_j : j = 1, \cdots, k\}$ (refer to §2.2) remains on this subspace for all forward and reverse time. Similarly, the (two-dimensional) subspace spanned by the real part of $\{v_j\}$ and the imaginary part of $\{v_j\}$, when v_j is a complex eigenvector, is invariant under the flow $\phi_t(\mathbf{u}) = e^{tA}\mathbf{u}$. A subspace formed by the span of eigenvectors is also known as an eigenspace. In short, any eigenspace of A is an invariant subspace of the flow. The subspaces spanned by the eigenvectors of A can be divided into three types (Guckenheimer and Holmes 1983, page 10):

$$\begin{aligned} &\text{the stable subspace, } E^s = span\{\mathbf{v}_1, \cdots, \mathbf{v}_{n_s}\}, \\ &\text{the unstable subspace, } E^u = span\{\mathbf{u}_1, \cdots, \mathbf{u}_{n_u}\}, \\ &\text{the centre subspace, } E^c = span\{\mathbf{w}_1, \cdots, \mathbf{w}_{n_c}\}, \end{aligned} \tag{1.19}$$

where $\mathbf{v}_1, \cdots, \mathbf{v}_{n_s}$ are the n_s eigenvectors whose eigenvalues have negative real parts, $\mathbf{u}_1, \cdots, \mathbf{u}_{n_u}$ are the n_u eigenvectors whose eigenvalues have positive real parts, and $\mathbf{w}_1, \cdots, \mathbf{w}_{n_c}$ are the n_c eigenvectors whose eigenvalues have zero real parts and $n_s + n_u + n_c = n$. The trajectories lying on E^s are characterised by exponential decay (either monotonic or oscillatory), those lying in E^u by exponential growth, and those lying in E^c by neither. The usefulness of splitting state space into these three invariant subspaces is that once a point in state space is specified relative to these subspaces, the qualitative behaviour of the trajectory of that point is immediately known.

The long term or *asymptotic behaviour* of a trajectory is the region (i.e. the subset) of state space approached asymptotically as $t \to \infty$ (Guckenheimer and Holmes 1983, page 3). The asymptotic behaviour of a trajectory may be a single point (i.e. a trajectory asymptotically approaches a point in state space as $t \to \infty$), or an infinity of points, in which case the asymptotic behaviour involves motion in state space. Any behaviour that is not asymptotic is known as *transient.* For example, consider the state portrait shown in figure 1.4(c). The asymptotic behaviour of the trajectory of any point that lies on the x_1 axis in figure 1.4(c) (which is a stable subspace) is the point $(0,0)$ in state space. The asymptotic behaviour of the trajectory of any point not on the x_1 axis is $(0, \pm\infty)$, and such a trajectory is said to be *unbounded.* A trajectory is said to be unbounded or unbounded in forward time, if any state space coordinate specifying the trajectory tends to infinity in forward time (i.e. $t \to \infty$).

A point $\bar{\mathbf{x}}$ is called an *equilibrium point* of (1.9) if $A\bar{\mathbf{x}} = 0$ (Hirsch and Smale 1974, page 180). The constant function $\mathbf{x}(t) = \bar{\mathbf{x}}$ is then a solution to (1.9). If the flow associated with (1.9) is considered, it is seen that $\phi_t(\bar{\mathbf{x}}) = \bar{\mathbf{x}}$ for all time. Since the flow is a constant at $\bar{\mathbf{x}}$, $\bar{\mathbf{x}}$ is called a *fixed* or *stationary point* of the flow (Hirsch and Smale 1974, page 181). The point $(0,0)$ in each of the state portraits shown in figure 1.4 is a fixed point.

A fixed point is called a *hyperbolic* or *non-degenerate fixed point* if no eigenvalue of A has a zero real part and if each eigenvalue is different (Guckenheimer and Holmes 1983, page 17). A fixed point is said to be *attracting* or *asymptotically stable* if a neighbourhood of $\bar{\mathbf{x}}$ can be found such that the asymptotic behaviour of the trajectory of every point within this neighbourhood is $\bar{\mathbf{x}}$ (Guckenheimer and Holmes 1983, page 3). A fixed point can only be attracting if all the eigenvalues of A have negative real parts (Percival and Richards 1982, Chapter 3). For a linear system, if an attracting fixed point exists, it is globally attracting (i.e. the asymptotic behaviour of the trajectory of every point on $\mathbf{R}^n$ is $\bar{\mathbf{x}}$). A fixed point is said to be *repelling* or *unstable* if it is not attracting (i.e. a fixed point is repelling if at least one of the eigenvalues of A has a positive real part). The fixed point in figure 1.4(a) is an attracting fixed point while the fixed points in figure 1.4(b) and (c) are repelling. Determining if a fixed point is an attracting fixed point or a repelling fixed point is called finding the stability of the fixed point. A linear system can have at most one fixed point. At fixed points invariant subspaces connect (refer to §2.3).

A fixed point is called a degenerate fixed point if one or more of the eigenvalues of A has a zero real part, or if two or more of the eigenvalues are equal. If the A matrix forming a degenerate fixed point is perturbed slightly, then in general a non-degenerate fixed point results. A degenerate fixed point is structurally unstable (refer to §2.3) since any small perturbation to the A matrix (i.e. a small perturbation of the system equations) destroys the degenerate fixed point, which is not the case for non-degenerate fixed points. Degenerate fixed points almost never arise in physical systems, since almost all matrices result in non-degenerate fixed points. Degenerate fixed points usually only arise if friction or losses are ignored. The standard matrix type corresponding to the state portrait shown in figure 1.3(c) (i.e. type 3 matrix) is structurally unstable. In linear systems, structurally stable bounded asymptotic behaviour can only be associated with attracting fixed points.

Real eigenvalues (and the real parts of complex eigenvalues) result in the velocity of motion of a trajectory towards or away from a fixed point being of the form $e^{\lambda t}$,

where $\hat{\lambda}$ is an eigenvalue (i.e. the rate of motion of a trajectory towards an attracting fixed point exponentially slows down as the fixed point is approached, and the rate of motion away from a repelling fixed point exponentially increases as the trajectory moves further away). Complex eigenvalues always occur in complex conjugate pairs, and cause circular or oscillatory motion of trajectories around fixed points (Hirsch and Smale 1974). A pair of purely imaginary eigenvalues cause circular motion around a fixed point in state space. Such fixed points are structurally unstable, and are often called centres. A pair of complex eigenvalues having a real and imaginary component have trajectories which spiral into or away from a fixed point. Such fixed points are called elliptic fixed points, to emphasis the spiralling and circular behaviour of trajectories in the neighbourhood of such fixed points. If trajectories in the neighbourhood of the fixed point spiral into the fixed point (i.e. if the real part of the complex eigenvalues are negative) then the fixed point is called an attracting elliptic fixed point. If the opposite happens the fixed point is called a repelling elliptic fixed point.

1.1.2 The Nonlinear System

A nonlinear system is any system which cannot be represented by (1.9). The behaviour of nonlinear systems are much less well understood in general than the behaviour of linear systems. Unlike linear systems, nonlinear systems can have asymptotic behaviour other than fixed points. The general nonlinear system is described by

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) \tag{1.20}$$

where $\mathbf{x}$ and $\mathbf{f}$ are vectors. A theorem by Hartman and Grobman (Guckenheimer and Holmes 1983, page 13) enables the behaviour near a fixed point $\bar{\mathbf{x}}$ (i.e. the stability) to be completely described by linearizing $\mathbf{f}(\mathbf{x})$ about $\bar{\mathbf{x}}$ (i.e. by taking the Taylor series expansion about $\bar{\mathbf{x}}$ and keeping only the first term) and using the methods of linear algebra. Linearizing $\mathbf{f}(\mathbf{x})$ about $\bar{\mathbf{x}}$ gives

$$\begin{aligned} \frac{d(\bar{\mathbf{x}}+\xi)}{dt} &= \mathbf{f}(\bar{\mathbf{x}}) + \mathbf{f}'(\bar{\mathbf{x}})\xi + \frac{\mathbf{f}''(\bar{\mathbf{x}})\xi^2}{2} + \cdots \\ \dot{\xi} &= \mathbf{f}'(\bar{\mathbf{x}})\xi + \frac{\mathbf{f}''(\bar{\mathbf{x}})\xi^2}{2} + \cdots \end{aligned} \tag{1.21}$$

and taking the first term only

$$\dot{\xi} = \mathbf{f}'(\bar{\mathbf{x}})\xi \tag{1.22}$$

where ξ is the distance (refer to §2.1) from $\bar{\mathbf{x}}$ and $\mathbf{f}' = [\delta f_i/\delta x_j]$ is the Jacobian matrix of first partial derivatives of the function $\mathbf{f}(\mathbf{x}) = (f_1(x_1, \cdots, x_n), \cdots, f_n(x_1, \cdots, x_n))^T$ (T denotes transpose). The linearized flow arising from (1.20) at a fixed point $\bar{\mathbf{x}}$ is obtained in the same way as for a linear system.

The Hartman and Grobman Theorem states that if $\mathbf{f}'(\bar{\mathbf{x}})$ has no eigenvalues with a zero real part, then there exists a homomorphism $\mathbf{h}$ (an invertible continuous map; refer to §2.3) defined on some neighbourhood $N_{\bar{x}}$ of $\bar{\mathbf{x}}$ taking trajectories of the nonlinear flow arising from (1.20), to those of the linearized flow arising from (1.22). The homomorphism preserves the sense (or direction) of trajectories. Two flows are said to be equivalent in qualitative terms, if there exists a homomorphism

taking the trajectories of one to the trajectories of the other. This theorem implies that the types of behaviour possible near fixed points in nonlinear systems are the same as the types of behaviour possible in linear systems. It turns out that the fixed points arising in nonlinear systems can be classified in exactly the same way as those arising in linear systems (Hirsch and Smale 1974).

In a one-dimensional nonlinear system, the only kind of bounded asymptotic behaviour possible which is structurally stable is an attracting fixed point. In $\mathbf{R}^2$ one additional type of bounded asymptotic behaviour is possible, called a *periodic orbit.* A periodic orbit is a trajectory $\mathbf{x}(t)$ for which there exists $0 < T < \infty$ such that $\mathbf{x}(t) = \mathbf{x}(t + T)$ for all t. The word 'orbit' is usually invoked, as opposed to 'trajectory' because the trajectory of a periodic orbit is a simple closed curve and the term orbit is more appropriate. A structurally stable periodic orbit is called a limit cycle (Hirsch and Smale 1974). The stability of a *limit cycle* is determined from Floquet multipliers or equivalently from the Poincaré map (refer to §1.1.4) (Guckenheimer and Holmes 1983, page 22). The same stability classification can be used for limit cycles (hyperbolic, attracting, repelling) as for fixed points.

The existence and uniqueness theorem for solution curves of ODEs states that every solution curve is unique, that is, the trajectory of every point in state space has a unique future and a unique past (Guckenheimer and Holmes 1983, page 15). The consequence of this is that two different solution curves cannot intersect in anyway since the point(s) at the intersection could have evolved from two different pasts and/or could evolve towards two different futures. For the same reason, a solution curve cannot cross or meet itself at a non-zero angle (i.e. transversally). However, a solution curve can form a simple closed curve (i.e. the only kind of trajectory that can cross or meet itself is a trajectory which forms a simple closed curve, e.g. a periodic orbit). Due to the existence and uniqueness theorem, the type of behaviour in one and two-dimensional state space is severely restricted and results in relatively uninteresting behaviour (i.e. fixed points and periodic orbits) (Hirsch and Smale 1974, Chapter 11).

Additional types of asymptotic behaviour are possible in $\mathbf{R}^3$. A generalisation of the two-dimensional periodic orbit is a trajectory confined to the surface of a torus (in general a distorted torus). This trajectory or orbit is formed from the Cartesian product (refer to §2.1) of two (or more) independent periodic orbits. Orbits formed from the Cartesian products of periodic orbits are called *quasi-periodic orbits.* Behaviour characterised by either a constant or the Cartesian product (possibly infinite) of periodic orbits is called regular behaviour. The only types of *regular behaviour* possible are characterised by fixed points, periodic orbits and quasi-periodic orbits.

Analogues of the stable and unstable invariant subspaces of the linear system exist for the nonlinear system. Instead of these invariant subspaces being straight lines or (hyper)planes they are smooth curves or (hyper)surfaces called *invariant manifolds* (refer to §2.1). Invariant manifolds are conveniently defined in two steps (Guckenheimer and Holmes 1983, page 13). First, local invariant manifolds are defined in the neighbourhood of a fixed point, since invariant manifolds are always present near, and always connect to each other at fixed points. Second, local invariant manifolds are extended globally by letting points on them flow either backward or forward in time. The local stable and unstable invariant manifolds of $\bar{\mathbf{x}}$, denoted by

$W^s_{loc}(\bar{\mathbf{x}})$, $W^u_{loc}(\bar{\mathbf{x}})$ respectively, are defined by

$$\begin{aligned} W^s_{loc}(\bar{\mathbf{x}}) &= \{\mathbf{x} \in N_{\bar{x}} : \phi_t(\mathbf{x}) \to \bar{\mathbf{x}} \text{ as } t \to \infty, \text{ and } \phi_t(\mathbf{x}) \in N_{\bar{x}} \ \forall t \geq 0\} \\ W^u_{loc}(\bar{\mathbf{x}}) &= \{\mathbf{x} \in N_{\bar{x}} : \phi_t(\mathbf{x}) \to \bar{\mathbf{x}} \text{ as } t \to -\infty, \text{ and } \phi_t(\mathbf{x}) \in N_{\bar{x}} \ \forall t \leq 0\} \end{aligned} \tag{1.23}$$

where $N_{\bar{x}} \subset \mathbf{R}^n$ is a neighbourhood of the fixed point $\bar{\mathbf{x}}$. The local invariant manifolds $W^s_{loc}(\bar{\mathbf{x}})$ and $W^u_{loc}(\bar{\mathbf{x}})$ provide nonlinear analogues of the (flat) stable and unstable subspaces E^s, E^u of the linear system. Suppose $\bar{\mathbf{x}}$ is a hyperbolic fixed point, then there exists local stable and unstable manifolds $W^s_{loc}(\bar{\mathbf{x}})$, $W^u_{loc}(\bar{\mathbf{x}})$, of the same dimensions n_s, n_u as those of the eigenspaces E^s, E^u of the linearized system, and tangent to E^s, E^u at $\bar{x}$. The local invariant manifolds $W^s_{loc}(\bar{x})$, $W^u_{loc}(\bar{\mathbf{x}})$, have global analogies $W^s(\bar{\mathbf{x}})$, $W^u(\bar{\mathbf{x}})$, obtained by letting points on $W^s_{loc}(\bar{\mathbf{x}})$ flow backwards in time and those on $W^u_{loc}(\bar{\mathbf{x}})$ flow forward in time, i.e.

$$\begin{aligned} W^s(\bar{\mathbf{x}}) &= \bigcup_{t \leq 0} \phi_t(W^s_{loc}(\bar{\mathbf{x}})), \\ W^u(\bar{\mathbf{x}}) &= \bigcup_{t \geq 0} \phi_t(W^u_{loc}(\bar{\mathbf{x}})), \end{aligned} \tag{1.24}$$

To ensure the uniqueness of every trajectory (i.e. that the trajectory of every point in state space has only one past and one future) the following rules concerning the intersection of invariant manifolds hold. Stable invariant manifolds emanating from different fixed points cannot intersect. Unstable invariant manifolds emanating from different fixed points cannot intersect. An invariant manifold cannot cross or meet itself. However, a stable and an unstable invariant manifold emanating from different fixed points or emanating from the same fixed point can intersect. The stable and unstable invariant manifolds emanating from a fixed point can intersect smoothly (i.e. coincide exactly). The trajectory of a point on such an intersection is called a *homoclinic trajectory* (Guckenheimer and Holmes 1983, page 22). The asymptotic behaviour of such a trajectory in both forward and reverse time (i.e. for $t \to \infty$ and $t \to -\infty$) is characterised by the same point, namely the fixed point. To emphasise the special nature of such a fixed point, it is called a *homoclinic fixed point.* A homoclinic trajectory forms a closed loop like a periodic orbit, but unlike a periodic orbit it cannot be traversed in a finite time. It is sometimes useful to think of the homoclinic trajectory as a limiting case of a periodic orbit. If the stable and unstable invariant manifolds emanating from different fixed points intersect smoothly (i.e. coincide exactly) then the trajectory of a point on such an intersection is called a *heteroclinic trajectory.* Systems which possess homoclinic or heteroclinic trajectories are structurally unstable.

When a system which possesses homoclinic (or heteroclinic) trajectories is perturbed (e.g. by the addition of losses, an external driving function, etc.), the stable and unstable invariant manifolds forming the homoclinic (or heteroclinic) trajectory separate. The resulting behaviour of the perturbed system depends on how this separation takes place. If the stable and unstable manifolds separate completely, regular behaviour results. However, if the stable and unstable manifolds separate in a twisted way intersecting transversely, complicated or apparently chaotic motion can occur. The transverse intersections are called *transverse homoclinic* (or heteroclinic) *points.* There is increasing evidence that many natural systems that appear to behave chaoticly, involve the appearance of homoclinic or heteroclinic points (Nicolis

1986, page 903). The goal of the following sections in this Chapter is to provide a qualitative explanation of why apparently chaotic dynamical behaviour can arise in the neighbourhood of homoclinic and heteroclinic points. Before launching off into this explanation, it is useful to consolidate many of the concepts introduced in §1.1.1 and §1.1.2, by an illustrative example. The rest of this subsection concentrates on examining the dynamics of a pendulum. This pendulum example is revisited in §1.1.10, and consolidates many of the concepts introduced in the subsections preceding §1.1.10.

The system equations characterising a frictionless pendulum are described by (1.5), and the corresponding state portrait is shown in figure 1.2. The fixed points $\bar{\mathbf{x}}$ of (1.5) are found by setting the velocity function to zero, i.e.

$$\begin{aligned} \dot{x}_1 &= \bar{x_2} = 0 \\ \dot{x}_2 &= -\sin \bar{x_1} = 0 \end{aligned} \tag{1.25}$$

which can be solved for $\bar{x}_1$ and $\bar{x}_2$,

$$\begin{aligned} \bar{x}_1 &= n\pi, \quad \text{for } n \text{ integer} \\ \bar{x}_2 &= 0 \end{aligned} \tag{1.26}$$

There exists an infinite number of fixed points. They lie along the x_1 axis in state space separated by a distance of π. The stability or the type of each fixed point is found by calculating the eigenvalues $\hat{\lambda}_i$ of the linearized system (i.e. the Jacobian matrix of the velocity function) about each fixed point. The Jacobian matrix is

$$\mathbf{f}'(\bar{\mathbf{x}}) = \begin{pmatrix} 0 & 1 \\ -\cos(n\pi) & 0 \end{pmatrix}, \quad \text{for } n \text{ integer}, \tag{1.27}$$

and the eigenvalues of $\mathbf{f}'(\bar{\mathbf{x}})$ are found by solving (*cf.* Hirsch and Smale 1974)

$$\begin{aligned} |\, \mathbf{f}'(\bar{\mathbf{x}}) - I\hat{\lambda}\,| &= 0 \\ \Rightarrow \begin{vmatrix} -\hat{\lambda} & 1 \\ -\cos n\pi & -\hat{\lambda} \end{vmatrix} &= 0 \\ \Rightarrow \hat{\lambda}^2 &= -\cos n\pi \end{aligned} \tag{1.28}$$

where I is the identity matrix, and $|\cdot|$ denotes the determinant. The eigenvectors of the Jacobian matrix define the invariant subspaces or eigenspaces of the linearized system. The invariant manifolds of the nonlinear system are tangent to and intersect the invariant subspaces of the linearized system at the fixed points. The eigenvectors of $\mathbf{f}'(\bar{\mathbf{x}})$ are found by solving (*cf.* Hirsch and Smale 1974)

$$\mathbf{f}'(\bar{\mathbf{x}})v_j = \hat{\lambda}_j v_j \tag{1.29}$$

The fixed points at the coordinates $\bar{\mathbf{x}} = (0, n\pi)$, for which n is an odd integer, denoted $\bar{\mathbf{x}}_{odd}$, are of a different type from the fixed points for which n is an even integer, denoted $\bar{\mathbf{x}}_{even}$. The eigenvalues and eigenvectors of the system linearized about $\bar{\mathbf{x}}_{odd}$ are

$$\begin{aligned} \hat{\lambda}_1 = 1 \text{ and } v_1 &= \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ \hat{\lambda}_2 = -1 \text{ and } v_2 &= \begin{pmatrix} 1 \\ -1 \end{pmatrix} \end{aligned} \tag{1.30}$$

and the eigenvalues and eigenvectors of the system linearized about $\bar{x}_{even}$ are

$$\begin{aligned} \hat{\lambda}_1 = j1 \text{ and } v_1 &= \begin{pmatrix} 1 \\ 0 \end{pmatrix} + j \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ \hat{\lambda}_2 = -j1 \text{ and } v_2 &= \begin{pmatrix} 1 \\ -1 \end{pmatrix} - j \begin{pmatrix} 0 \\ 1 \end{pmatrix} \end{aligned} \tag{1.31}$$

The $\bar{\mathbf{x}}_{odd}$ form hyperbolic fixed points, while the $\bar{\mathbf{x}}_{even}$ form elliptic fixed points. The eigenvectors of the hyperbolic fixed points form stable and unstable eigenspaces, while the eigenvectors of the elliptic fixed points form two-dimensional centre eigenspaces (i.e. the elliptic fixed points form centres). The elliptic fixed points characterise the pendulum at rest (i.e. sitting vertically downwards). If the pendulum is disturbed slightly, it oscillates about its rest position forever (since the pendulum is frictionless), forming periodic orbits in state space (note that the elliptic fixed points are structurally unstable). The hyperbolic fixed points characterise the unstable position where the pendulum is standing vertically upwards. If the pendulum is perturbed slightly to one side when in this position, the pendulum swings away from the vertically upward position, and reapproaches the vertically upward position from the other side (i.e. does a complete rotation). This rotation is characterised by the heteroclinic trajectory, since neighbouring hyperbolic fixed points are asymptotically approached for $t \to \infty$, and $t \to -\infty$. In state space this means the stable and unstable manifolds of neighbouring hyperbolic fixed points coincide exactly.

Heteroclinic trajectories separate state space into distinct regions. The dynamical motion in each region is confined to that region. Curves or (hyper)surfaces that separate state space into distinct regions in this way are called *separatrixes*. Homoclinic and heteroclinic trajectories always form separatrixes, but not all separatrixes are homoclinic or heteroclinic trajectories.

In general the flow of individual trajectories of nonlinear systems cannot generally be evaluated. However, for this pendulum example, the heteroclinic trajectory can be solved. To do this it is necessary to examine the pendulum in terms of energy. Since the pendulum is frictionless, the total energy H of the system remains constant. The total energy of the pendulum is described by (1.8), from which the heteroclinic trajectory can be calculated. The heteroclinic trajectory is the trajectory that results when the pendulum has just sufficient energy to swing and stand vertically upwards. When in this position, the entire energy of the system is potential. The total energy of the system equals the maximum potential energy, which is unity. To calculate the heteroclinic trajectory, set $H = 1$ in (1.8) and solve for x_2 in terms of x_1

$$\begin{aligned} 1 &= \frac{1}{2}x_2^2 - \cos x_1 \\ x_2 &= \pm 2 \cos \frac{x_1}{2} \end{aligned} \tag{1.32}$$

The stable and unstable manifolds which coincide to form the heteroclinic trajectory, separate when friction is added to the pendulum. When friction is added the hyperbolic fixed points remain hyperbolic, but the non-attracting non-repelling elliptic fixed points (i.e. the centres) become attracting fixed points. Friction is a force which is proportional to the angular velocity $l\dot{\theta}$ of the pendulum. The restoring force becomes $F = -\varepsilon\dot{\theta} - \sin\theta$ where $\varepsilon\dot{\theta}$ represents the friction. The angular acceleration

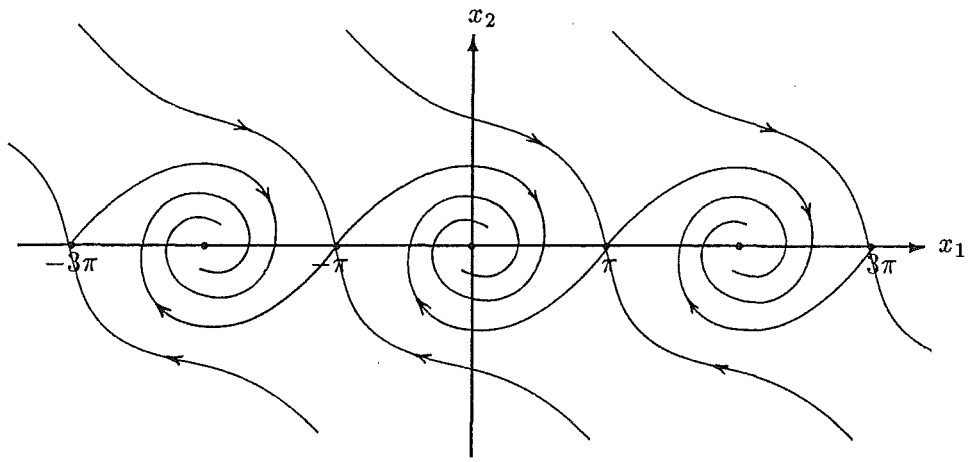


Figure 1.5: State portrait of the pendulum with friction, for $0 < \varepsilon < 2$. The $\bar{x}_{odd}$ form hyperbolic fixed points and the $\bar{x}_{even}$ form attracting elliptic fixed points. The stable and unstable invariant manifolds of the hyperbolic fixed points do not coincide, as is the case for the frictionless pendulum.

is

$$\ddot{\theta} = -\varepsilon\dot{\theta} - \sin\theta \tag{1.33}$$

which written as a system of first order ODEs, becomes

$$\begin{aligned} \dot{x_1} &= x_2 \\ \dot{x_2} &= -\varepsilon x_2 - \sin x_1 \end{aligned} \tag{1.34}$$

The fixed points $\bar{\mathbf{x}}$ occur at the same locations as before. The stabilities of the fixed points are found by calculating the eigenvalues of the Jacobian. The eigenvalues of $\bar{\mathbf{x}}_{odd}$ are

$$\begin{aligned} \hat{\lambda}_1 &= -\frac{\varepsilon}{2} + \frac{\sqrt{\varepsilon^2+4}}{2} \\ \hat{\lambda}_2 &= -\frac{\varepsilon}{2} - \frac{\sqrt{\varepsilon^2+4}}{2} \end{aligned} \tag{1.35}$$

For any value of $\varepsilon > 0$, $\hat{\lambda}_1$ is positive and $\hat{\lambda}_2$ is negative, thus $\bar{\mathbf{x}}_{odd}$ remain hyperbolic. The eigenvalues of $\bar{\mathbf{x}}_{even}$ are

$$\begin{aligned} \hat{\lambda}_1 &= -\frac{\varepsilon}{2} + \frac{\sqrt{\varepsilon^2-4}}{2} \\ \hat{\lambda}_2 &= -\frac{\varepsilon}{2} - \frac{\sqrt{\varepsilon^2-4}}{2} \end{aligned} \tag{1.36}$$

For $0 < \varepsilon < 2$, $\hat{\lambda}_1 = -\alpha + j\beta$, $\hat{\lambda}_2 = -\alpha - j\beta$ which is an attracting elliptical fixed point. For $\varepsilon > 2$, $\hat{\lambda}_1$ and $\hat{\lambda}_2$ are negative and distinct (i.e. a non-spiralling attracting fixed point). The state portrait for $0 < \varepsilon < 2$ is shown in figure 1.5. The stable and unstable invariant manifolds of the hyperbolic fixed points do not coincide. The unstable manifold spirals around the elliptic fixed points, and the stable manifold tend to infinity. Note that the stable manifolds of the hyperbolic fixed points form a separatrix.

1.1.3 Attractors

The concept of an attractor involves the behaviour (i.e. a subset of points in $\mathbf{R}^n$) a dissipative dynamical system evolves towards asymptotically as $t \to \infty$. More specifically, if $\phi_t(\mathbf{x})$ is a flow defined on $\mathbf{R}^n$, there almost always exist subsets Λ of $\mathbf{R}^n$ which attract neighbouring points $\mathbf{x}$. That is, $\phi_t(\mathbf{x})$ tends to Λ when $t \to \infty$. The study of attractors in dynamical systems has become increasingly important ever since Ruelle and Takens (1971) introduced the 'strange attractor' as a possible explanation for fluid turbulence. According to Ruelle (1981, page 138), the concept of an attractor usually encompass the following notions:

- Invariance: points within Λ remain within Λ for all time.
- Attractivity: points close to Λ tend to Λ as $t \to \infty$.
- Closedness or Compactness: the set Λ is finite in size and has a boundary.
- Irreducibility: if Λ consists of disjoint invariant pieces, one would like to consider each piece separately removing irrelevant points.
- Stability: points close to Λ remain close to Λ.
- A positive probability that a randomly chosen point in some neighbourhood of Λ is attracted to Λ.

Consider as an example the following dynamical system

$$\begin{aligned} \dot{x}_1 &= x_1 - x_1^3 \\ \dot{x}_2 &= x_2 \end{aligned} \tag{1.37}$$

the state portrait for (1.37) is shown in figure 1.6. There are two attracting fixed points at $(x_1, x_2) = (\pm 1, 0)$ and one repelling fixed point at the origin. Suppose some neighbourhood N contain these three fixed points. The set N is mapped into itself by the flow $\phi_t(N) \subset N$, and asymptotically contracts onto the closed interval [-1,1] (denoted A) as $t \to \infty$. The set A forms a closed invariant set (i.e. $\phi_t(A) = A$). An adequate definition of an attractor for a flow $\phi_t(\mathbf{x})$ might appear to be the following: a closed invariant set $A \subset \mathbf{R}^n$ is an attractor if there is some neighbourhood N of A such that for all $\mathbf{x} \in N$, $\phi_t(\mathbf{x}) \in N$ for all $t \geq 0$, and $\phi_t(\mathbf{x}) \to A$ as $t \to \infty$ (Guckenheimer and Holmes 1983, page 256). Invoking this definition of attractor for the dynamical system (1.37) results in the attractor being the closed interval [-1,1].

There are two noteworthy aspects of the example described in the previous paragraph (Guckenheimer and Holmes 1983, page 256). First, almost all points in N are attracted to the fixed points $(\pm 1, 0)$. In a physical system it is the fixed points at $(\pm 1, 0)$ which are important, since almost all points in N evolve towards these points $(\pm 1, 0)$ – it is essential that any definition of an attractor should make this distinction. Second, the attractor is a stable attractor in the sense that all points in N remain within N for all time (i.e. $\phi_t(N) \subset N$ for $t \geq 0$). This excludes unstable attractors in which points arbitrarily close to an attractor wander far away before converging back towards the attractor. Because of these aspects, the attractor definition of the previous paragraph is usually inadequate, but it is still an important concept and is termed an *attracting set* (Guckenheimer and Holmes 1983, page 34).

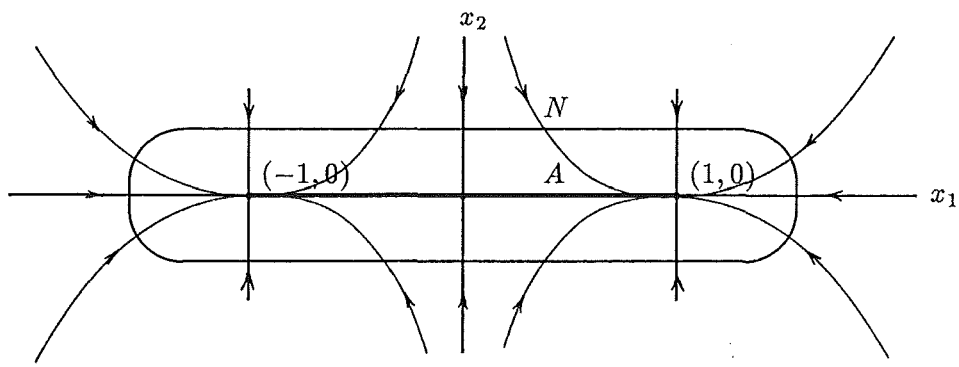


Figure 1.6: State portrait for the equation given by (1.37). There are two attracting fixed points at $(x_1, x_2) = (\pm 1, 0)$ and a repelling fixed point at the origin. The neighbourhood N contains the three fixed points.

The word attractor was first coined by Coddington and Levinson (1955, page 376) who applied it to a fixed point of a flow, but attractors consisting of more than one point were first studied by Auslander *et al.* (1964). He considered the unstable case in which a trajectory starting arbitrarily close to an attractor may wander far away before converging back towards the attractor (La Salle 1976). This has not been widely accepted and most authors have required some form of stability as part of their definitions. There are many definitions of attractor in the literature. Different attractor definitions can be found in Smale (1967, page 786) , Abraham and Marsden (1978, page 517) , Guckenheimer and Holmes (1983, page 36) , Guckenheimer and Holmes (1983, page 256) , Lichtenberg and Lieberman (1983, page 384) and Percival and Richards (1982, page 26) . A discussion on attractor definitions is given by Ruelle (1981) and Milnor (1985) .

In this thesis an attractor is defined to be the set of points Λ to which almost all points in any arbitrary small neighbourhood N of Λ evolve under the flow $\phi_t(\mathbf{x})$ as $t \to \infty$, such that $\phi_t(N) \subset N \ \forall t > 0$. This definition has two useful practical consequences. First, since almost all points in a neighbourhood N of Λ are attracted to Λ (i.e. any point chosen at random from N evolves towards Λ), such an attractor can always be detected practically (i.e. through a computer simulation or laboratory experiment). Second, if the trajectory of points initially close to Λ wander far away from Λ before returning to Λ, then in many practical situations Λ can be considered unstable. The stability requirement (i.e. $\phi_t(N) \subset N \ \forall t < 0$) ensures that points close to Λ remain close.

There are three types of regular attractor: fixed points, periodic orbits, and quasi-periodic orbits. All other attractors are generally termed strange by many authors (Hirsch 1984, page 30). However, *strange attractors* can further be split into two categories: 1) those that exhibit sensitive dependence on initial conditions and 2) those which have a non-integer spatial dimension in state space. It appears that when the term strange attractor is used in the literature, the first category is usually meant, but often this distinction is not made clear. This is possibly because there are not many examples of attractors that exhibit sensitive dependence on initial

conditions and also have an integer spatial dimension. However, there are exceptions in the literature, Grebogi *et al.* (1984) refers to the first category of attractor as a *chaotic attractor*, and the second category as a strange attractor. In this thesis the term *strange attractor* refers to attractors which exhibit sensitive dependence on initial conditions, emphasising that the notion of strangeness refers to the flow on the attractor and not its geometry in state space.

An example of an attractor which does not exhibit sensitive dependence on initial conditions but has a non-integer spatial dimension in state space, is the Feigenbaum attractor (refer to §1.1.6 and Eckmann and Ruelle 1985, page 625). An attractor which exhibits sensitive dependence on initial conditions (i.e. a strange attractor) can have an integer or non-integer spatial dimension in state space. Most known strange attractors have non-integer dimensions. An example of a strange attractor with an integer dimension is the hyperbolic toral automorphism (*cf.* Devaney 1987).

1.1.4 Poincaré Maps

When determining if a periodic orbit of a differential equation is attracting or repelling (i.e. determining the stability of the periodic orbit), one traditionally invokes what are called the characteristic or Floquet multipliers (*cf.* Tomita 1982, page 118). In this thesis an equivalent, but more geometrical viewpoint, known as the Poincaré map (described below) is taken.

Let γ be a periodic orbit of period T of some flow $\phi_t(\mathbf{x})$ in $\mathbf{R}^n$ arising from an ODE. Consider a cross section $\Sigma \subset \mathbf{R}^n$ of dimension $n-1$ of the flow. The surface Σ is chosen to intersect the flow transversely. This occurs if the dot product $\phi_t(\mathbf{x}) \cdot n(\mathbf{x}) \neq 0$ for all $\mathbf{x} \in \Sigma$, where $n(\mathbf{x})$ is the unit normal to Σ at $\mathbf{x}$. Denote the point where γ intersects Σ by $\mathbf{p}$, and let $N_p \subseteq \Sigma$ be some neighbourhood of $\mathbf{p}$. Then the Poincaré map $P : N_p \to \Sigma$ is defined for a point $\mathbf{q} \in N_p$ by

$$P(\mathbf{q}) = \phi_\tau(\mathbf{q}) \tag{1.38}$$

where $\tau = \tau(\mathbf{q})$ is the time taken for the orbit $\phi_t(\mathbf{q})$ at $\mathbf{q}$ to first return to Σ. Note that τ generally depends upon $\mathbf{q}$ and need not be equal to T, the period of γ (Guckenheimer and Holmes 1983, page 23). However, $\tau \to T$ as $\mathbf{q} \to \mathbf{p}$.

The point $\mathbf{p}$ is a fixed point of the map P, and the stability of $\mathbf{p}$ for P reflects the stability of γ for the flow ϕ_t (Guckenheimer and Holmes 1983, page 23). In particular, if $\mathbf{p}$ is a hyperbolic fixed point, and $P'(\mathbf{p})$ is the Jacobian matrix of P about $\mathbf{p}$ (i.e. the linearized map about $\mathbf{p}$ as discussed in §1.1.2) then $P'(\mathbf{p})$ has n_s eigenvalues with modulus less than unity and has n_u eigenvalues with modulus greater than unity, where $n_s + n_u = n - 2$ (i.e. the dimension of $W^s(\mathbf{p}) = n_s$, and the dimension of $W^u(\mathbf{p}) = n_u$ for the map P). Since the trajectories of P lying in $W^s(\mathbf{p})$ and $W^u(\mathbf{p})$ are formed by intersections of trajectories of $\phi_t(\cdot)$ with Σ, the dimensions of $W^s(\gamma)$ and $W^u(\gamma)$ are each one greater than those for the map (Guckenheimer and Holmes 1983, page 13).

The construction of a Poincaré map for a non-autonomous system is in general simpler than for an autonomous system. Recall that the velocity function of a non-autonomous system depends explicitly on $\mathbf{x}$ and t (i.e. $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t)$). If there exists a $T > 0$ such that $\mathbf{f}(\mathbf{x}, t) = \mathbf{f}(\mathbf{x}, t+T)$ the system is said to be *time-periodic* with period T. An n^{th}-order time-periodic non-autonomous system can always be converted to

an $(n+1)^{th}$ order autonomous system by appending an extra equation $\theta = 2\pi t/T$ (Guckenheimer and Holmes 1983). The autonomous system is given by

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{f}(\mathbf{x}, \theta T/2\pi) \\ \dot{\theta} &= 2\pi/T \end{aligned} \tag{1.39}$$

Since the $\mathbf{f}$ in (1.39) is time-periodic with period T, the system (1.39) is periodic in θ with period 2π. It is therefore possible to transform (1.39) from the $\mathbf{R}^{n+1}$ Euclidean state space to the cylindrical state space $\mathbf{R}^n \times \mathbf{S}$. The result of transforming (1.39) into cylindrical state space is

$$\begin{aligned} \mathbf{x}(t) &= \phi_t(\mathbf{u}, t_0) \\ \theta(t) &= (2\pi t/T) \mod(2\pi) \end{aligned} \tag{1.40}$$

where $\mod(2\pi)$ means that θ is restricted to the interval $[0, 2\pi)$. After this transformation, the results from autonomous systems can be applied to the time-periodic non-autonomous case. The Poincaré map for an n^{th}-order non-autonomous system is a n-dimensional hyper-plane Σ in $\mathbf{R}^n \times \mathbf{S}$ defined by

$$\Sigma = \{(\mathbf{x}, \theta) : \mathbf{x} \in \mathbf{R}^n, \theta = \theta_0\} \tag{1.41}$$

where θ_0 is arbitrary. Every T seconds the trajectory of the non-autonomous system intersects Σ. Thus this Poincaré map is defined by $P(\mathbf{q}) = \phi_T(\mathbf{q}, \theta_0)$, and can be considered similar to the action of a stroboscope flashing with period T (Guckenheimer and Holmes 1983, page 26).

1.1.5 Discrete Maps

A discrete dynamical system is a system that generates a sequence of elements or points characterised by numbers or vectors. A (semi-)deterministic discrete dynamical system can always be arranged so that the next element of the sequence is a function (map) of the previous element of the sequence only. The equations that describe discrete dynamical systems are called *discrete maps* or *difference equations.* A discrete map is a system of the form

$$\mathbf{x}_{n+1} = \mathbf{f}(\mathbf{x}_n), \quad n = 0, 1, \cdots \tag{1.42}$$

where the next element of the sequence $\mathbf{x}_{n+1}$ is dependant on the present element $\mathbf{x}_n$ only (*cf.* May 1976; Collet and Eckmann 1980; Devaney 1987). Discrete maps can arise directly from physical phenomenon (e.g. population biology) or by sampling a continuous dynamical process (May 1976; Guckenheimer *et al.* 1977).

There are several ways a trajectory of a continuous dynamical system could be sampled in order to create a discrete dynamical system. If the flow of a dynamical system is sampled uniformly in time at instants spaced by τ, and recalling that the group properties (1.2) hold for the flow $\phi_t(\cdot)$, then a discrete map results

$$\begin{aligned} \mathbf{x}_{n+1} &= \phi_\tau(\mathbf{x}_n) = \mathbf{f}(\mathbf{x}_n) \\ \mathbf{x}_{n+2} &= \phi_\tau(\phi_\tau(\mathbf{x}_n)) = \mathbf{f}(\mathbf{f}(\mathbf{x}_n)) \end{aligned} \tag{1.43}$$

Since the flow $\phi_t 9\cdot)$ is smooth, $\mathbf{f}$ is a smooth map and is called a diffeomorphism (refer to §2.3). A more important way to sample the trajectory of a continuous dynamical

system is via the Poincaré map (refer to §1.1.4). The Poincaré map reduces the dimensionality of the continuous dynamical system by one but preserves the motion of trajectories in the other dimensions. The Poincaré map is invertible since it is possible in integrate an ODE forward or backward in time (since ODEs model deterministic processes), and results in the Poincaré map being a diffeomorphism. This is the main reason why the study of discrete maps are important, since discrete maps (in particular diffeomorphisms) are the simplest way to study dynamical systems. Smale (1967 page 747) was the first to really promote this idea and comments " ...there is a ...more important reason for studying the diffeomorphism problem (besides its great natural beauty). That is, the same phenomena and problems of the qualitative theory of ordinary differential equations are present in there simplest form in the diffeomorphism problem. Having first found theorems in the diffeomorphism case, it is usually a secondary task to translate the results back into the differential equations framework."

A trajectory of a discrete map is a sequence of points $\{\mathbf{x}_i : i = 0, 1, \cdots\}$. Any initial point $\mathbf{x}_0$ generates a unique trajectory. Stable, unstable and centre invariant subspaces for a linear map, or stable and unstable invariant manifolds for a non-linear map, can be defined as for continuous flows. The same types of asymptotic behaviour occur in discrete systems as occur in continuous systems. However, flows and maps differ crucially in that while the trajectory $\phi_t(\cdot)$ of a flow is a continuous curve in $\mathbf{R}^n$, the trajectory of a discrete map is a sequence of points in $\mathbf{R}^n$. The invariant manifolds of continuous flows are composed of the unions of continuous trajectories, those of maps are unions of discrete trajectory points. The complexity of behaviour of continuous dynamical systems is severely restricted in one and two dimensions due to the requirement that trajectories cannot meet or intersect. In discrete dynamical systems trajectories are a set of points (i.e. not continuous curves) so there are no restraints on the complexity of behaviour in low dimensional discrete systems. Seemingly stochastic or chaotic behaviour can occur in one-dimensional discrete dynamical systems.

1.1.6 One-Dimensional Maps

This subsection discusses three general aspects of one-dimensional maps. The first two aspects concern two theorems unique to one-dimensional maps (i.e. the theorems do not carry over into n dimensions), while the third aspect concerns the properties of a one-dimensional map characterised by a single parameter.

Consider the following ordering, named after Sarkovskii (*cf.* Kloeden and Mees 1985, page 705; Devaney 1987, page 60), of the natural numbers

$$3, 5, 7, \cdots 2 \times 3, 2 \times 5, \cdots 2^2 \times 3, 2^2 \times 5, \cdots 2^3, 2^2, 2 \tag{1.44}$$

Sarkovskii's theorem states that, if a one-dimensional continuous mapping $f : \mathbf{R} \to \mathbf{R}$ has a periodic orbit of period k, f contains all periodic orbits of periods greater than k in the Sarkovskii ordering of the natural numbers (e.g. if f contains a periodic orbit of period 2^3, then f also contains periodic orbits with periods of $2^2, 2$). However, these periodic orbits may not be attracting. This theorem implies that if $f : \mathbf{R} \to \mathbf{R}$ has a periodic orbit whose period is not a power of 2, then $f : \mathbf{R} \to \mathbf{R}$ necessarily has infinitely many different periodic orbits. Conversely, if $f : \mathbf{R} \to \mathbf{R}$ has only finitely many periodic orbits they all necessarily have periods which are powers of 2.

The existence of a period 3 orbit implies the existence of orbits of all other periods. Sarkovskii's theorem says further, that if a period 3 orbit exists, then there exist trajectories which are seemingly stochastic or chaotic. Sarkovskii's theorem was unknown in the West before its rediscovery by Li and Yorke (1975).

Allwright (1978) and Singer (1978) introduced into dynamics a property of one-dimensional maps which invokes the Schwarzian derivative. The Schwarzian derivative of a function $f(x)$ at the point x is defined to be

$$Sf(x) = \frac{f'''(x)}{f'(x)} - \frac{3}{2}\left(\frac{f''(x)}{f'(x)}\right)^2 \tag{1.45}$$

The Schwarzian derivative is invoked to establish the upper bound on the number of attracting periodic orbits a map may have, and to infer that certain maps have an interval on which the map has a seemingly stochastic or chaotic trajectory (*cf.* Kloeden and Mees 1985, page 709; Devaney 1987, page 98) If $Sf < 0$ over the domain of the map, and f has n critical points (i.e. f has n, x values where $f'(x) = 0$), then f can at most have $n+2$ attracting periodic orbits. Singer (1978) and Guckenheimer (1979) have provided a comprehensive analysis of such maps and their possible modes of behaviour.

The third aspect considered in this subsection concerns one-dimensional maps characterised by a single parameter (i.e. maps of the form $x_{n+1} = gf(x_n)$, where g is a real number parameter). The dynamics of such a map depends on the value of the parameter g. Such a map can arise, for example, in population biology (*cf.* May 1976; Guckenheimer *et al.* 1977). Consider, as an example, the population of a single species with non-overlapping breeding seasons. The population of the species in the $(n+1)^{th}$ season (denoted by P_{n+1}), depends on: the population of the species in the n^{th} season (denoted by P_n), the birth rate b, and on the death rate d of the species. The population in the $(n+1)^{th}$ season is given by (May 1976)

$$\begin{aligned} P_{n+1} &= bP_n - dP_n \\ &= (b-d)P_n \end{aligned} \tag{1.46}$$

Depending on the values of b and d there are two possibilities for the dynamics of (1.46): unchecked growth or extinction of the species. However, more realisticly the death rate generally increases as the population increases. This occurs because of overcrowding, lack of resources, etc. If the death rate increases in proportion to the population (i.e. $d \propto P_n$) (1.46) becomes

$$P_{n+1} = bP_n - \hat{d}P_n^2 \tag{1.47}$$

where $\hat{d} = d/P_n$. After the variable transformation, $x_n = \hat{d}/bP_n$, $x_{n+1} = \hat{d}/bP_{n+1}$ (1.47) becomes

$$x_{n+1} = gx_n(1-x_n) = f(x_n) \tag{1.48}$$

where $g = b$. Equation (1.48) is called the *logistic map* or *canonical equation.* It might be expected that the dynamics of the logistic map is similar to the linear equation (1.46). However, the logistic map can lead to the most complicated dynamics imaginable, seemingly stochastic or chaotic behaviour. An influential review

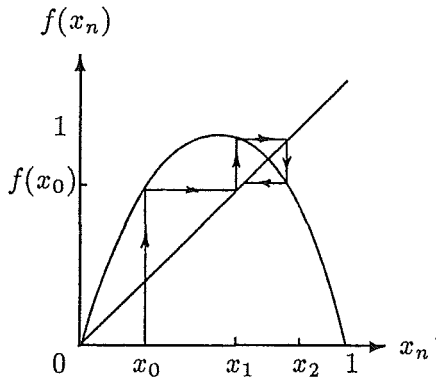


Figure 1.7: Diagrammatic illustration of the fixed point iteration given by equation (1.48). The initial condition x_0 is projected vertically to intersect $f(x_0)$, which gives the value of x_1. This intersection point is projected horizontally to intersect the $f(x) = x$ line. The intersection with the $f(x) = x$ line is projected vertically to intersect $f(x_1)$, which gives the value of x_2. This procedure if continued generates the entire trajectory of x_0.

paper by May (1976) was one of the first to emphasis that the notion of simple equations leading to simple behaviour can be wrong.

Equation (1.48) can be viewed as a fixed point iteration (May 1976; Collet and Eckmann 1980). The graphical representation of a fixed point iteration is shown in figure 1.7. If $0 < g < 1$ zero is an attracting fixed point (i.e. $\mid f'(0) \mid < 0$), for any starting point in the domain $0 < x_0 < 1$ (the initial condition). It is hereafter assumed that $x_0 \in [0,1]$. For $1 < g < 3$ zero is a repelling fixed point (i.e. $\mid f'(0) \mid > 1$ for $g > 1$), but a new attracting fixed point is spawned with the value $(g-1)/g$ (i.e. $\mid f'((g-1)/g) \mid < 1$ for $1 < g < 3$). Interesting behaviour begins for $g > 3$. If g is increased slightly above 3, the fixed point at $(g-1)/g$ becomes unstable (i.e. $\mid f'((g-1)/g) \mid > 1$ for $g > 3$). At the value of g for which the fixed point becomes unstable, two stable fixed points are spawned via a period doubling bifurcation (a bifurcation means a division in two, a splitting apart, a change, *cf.* Collet and Eckmann 1980; Mees 1983). These two new stable fixed points are special in that they form a periodic orbit of period two (i.e. $\bar{x}_2 = f(\bar{x}_1), \bar{x}_1 = f(\bar{x}_2)$ or $\bar{x}_2 = f^2(\bar{x}_2), \bar{x}_1 = f^2(\bar{x}_1)$), and is attracting for $3 < g < 3.25$ (i.e. $\mid f'(f'(\bar{x}_1)) \mid < 1$ for $3 < g < 3.25$). If g is increased slightly above 3.25, these two stable fixed points become unstable, and each spawns another two fixed points, to create a period four periodic orbit. Period doubling bifurcations continue as g is increased. The interval of g for which each periodic orbit is stable for, decreases as the period increases. A limiting value for g is reached for which an infinite number of period doubling bifurcations has taken place, this value of g is denoted g_∞.

At the limiting value $g_\infty = 3.57\cdots$ a special attractor known as the Feigenbaum attractor forms (Collet and Eckmann 1980; Eckmann and Ruelle 1985). Some of the properties of this attractor were discussed in §1.1.3. Interspersed with this attractor and arbitrarily close to it are repelling periodic orbits of periods 2^n for all n (Eckmann and Ruelle 1985, page 625). The interval $g_\infty < g < 4$ is termed the *chaotic* or *chaos region.* In this region there are trajectories that are seemingly stochastic or chaotic.

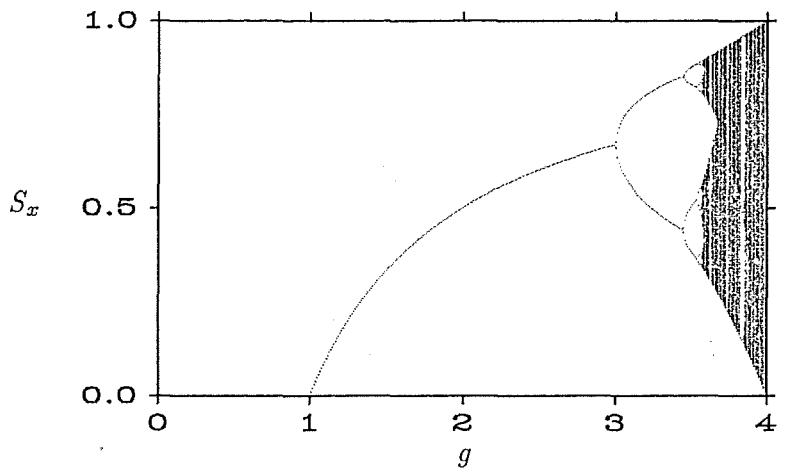


Figure 1.8: Bifurcation diagram of (1.48). The vertical axis represents the allowed region of state space for (1.48) (i.e. the interval $[0,1]$), and the horizontal axis represents the value of the parameter g (which can take on any value in the interval $[0,4]$). The asymptotic behaviour of (1.48) corresponding to the parameter value g specified on the horizontal axis is plotted vertically.

This region is also interspersed with intervals of g having attracting periodic orbits of all odd periods. The period doubling process that precedes the chaotic region in the logistic map is known as Feigenbaum's period doubling scenario preceding chaos (*cf.* §1.1.11). Feigenbaum (1978,1979,1980) has shown that for any map that has a negative Schwarzian derivative over the domain of the map, the same qualitative behaviour results as the parameter value for the map is increased, namely period doubling bifurcation culminating in a chaotic region.

A *bifurcation diagram* is a diagram which plots the system's asymptotic behaviour (or a representation of the asymptotic behaviour) against a system parameter (*cf.* Mees 1983). A bifurcation diagram for the logistic map (1.48) is shown in figure 1.8. The vertical axis shows the allowed region of state space for (1.48) (i.e. the interval $[0,1]$), and the horizontal axis shows the value of the parameter g (which can take on any value in the interval $[0,4]$). The asymptotic behaviour of (1.48) is approximated here by the sequence $S_x = \{x_i : i = n, \cdots, n+1000\}$ where n is made large enough so that initial transients are effectively negligible. The asymptotic behaviour (i.e. S_x) of (1.48) corresponding to the parameter value g specified on the horizontal axis is plotted vertically in figure 1.8. The branching exhibited by the bifurcation diagram (which is reminiscent to the branching of a tree) as g is increased from 3 to g_∞ represents the period doubling bifurcations preceding chaos. For $g > g_\infty$, the trajectories of (1.48) seemingly cover an entire region of the bifurcation diagram. This is the chaotic region. It also contains an infinite number of attracting periodic orbits with odd periods (a period three orbit can clearly be seen in the chaotic region of figure 1.8).

Maps which exhibit chaotic behaviour are here termed *chaotic maps*, while maps which do not are here termed *non-chaotic maps.* Any non-chaotic map can be perturbed by an arbitrarily small scaled down version of a chaotic map (e.g. the logistic

map (1.48) for $g = 4$) positioned at one of its equilibrium points, to produce a chaotic map (*cf.* Kloeden and Mees 1985, page 707). Such chaotic maps can be made arbitrarily close (in the sence of the C^0-distance; refer to §2.3) to the non-chaotic map. Thus, chaotic maps form a dense subset of the set of all maps (a subset is dense provided its closure is the entire set; refer to §2.1). Arbitrary perturbations do not cause major observable changes in the behaviour of a non-chaotic map. Kloeden (1976) comments that the perturbed trajectories (the trajectories of the chaotic map) might appear just as a smudge about the well behaved unperturbed trajectories (the trajectories of the non-chaotic map).

1.1.7 Symbolic Dynamics

Symbolic dynamics is a procedure for transforming the trajectories of a discrete dynamical system into a sequence of symbols (*cf.* Guckenheimer *et al.* 1977; Alekseev and Yakobson 1981; Guckenheimer and Holmes 1983; Devaney 1987). This enables the analysis of the discrete dynamics to be transformed into a problem of determining the combination of symbols in a symbol sequence, which is generally easier to analyse.

Each sequence of symbols defines a point in a sequence space Σ, and represents the trajectory of a point in the state space of the discrete dynamical system. Let there be n symbols consisting of the alphabet $0, 1, \cdots, n-1$. The sequence space defined by sequences of these n symbols is denoted Σ_n. The sequence space Σ_n represents the set of all symbol sequences, thus

$$\Sigma_n = \{S_s = (s_0, s_1, s_2, \cdots) : s_j \in \{0, \cdots, n-1\}\} \tag{1.49}$$

where S_s denotes a symbol sequence and s_j denotes the j^{th} member of the sequence.

The dynamics of the discrete dynamical system are characterised by what is termed the *shift map*. The shift map $\sigma : \Sigma_n \to \Sigma_n$ is defined as $\sigma(s_0, s_1, \cdots) = (s_1, s_2, \cdots)$, it removes the first entry in a sequence and shifts all other entries one place to the left. Devaney (1987 page 39) shows that it is possible to define a metric (refer to §2.1) on Σ_n which can be invoked to prove that the shift map forms a continuous mapping (refer to §2.3).

The dynamics of the shift map σ are well understood. Periodic orbits of σ correspond to points in Σ_n having repeating subsequences (i.e. sequences of the form $S_s = \{s_1, \cdots s_i, s_1, \cdots s_i, \cdots\}$). The points corresponding to repeating subsequences in Σ_n form a dense subset of Σ_n (recall that a subset is dense in Σ_n provided its closure is the entire Σ_n; refer to §2.1). However, not all sequences in Σ_n are repeating, in fact non-repeating sequences outnumber the repeating sequences (Devaney 1987, page 39). Moreover, there are points in Σ_n whose trajectory (under the shift map σ) come arbitrarily close to every point in Σ_n (i.e. such a trajectory forms a dense subset of Σ_n, or it is sometimes stated that such a trajectory is dense or winds densely around Σ_n).

To illustrate symbolic dynamics, consider as an example, the logistic map

$$x_{n+1} = f(x_n) = g x_n (1 - x_n) \quad \text{for} \quad g > 2 + \sqrt{5} \tag{1.50}$$

defined on the interval $[0, 1]$. Figure 1.9 illustrates the fixed point iteration given by (1.50). After one iteration, the interval labelled A in figure 1.9 leaves the interval $[0, 1]$ and the intervals labelled I_0 and I_1 each expand to cover the entire interval $[0, 1]$.

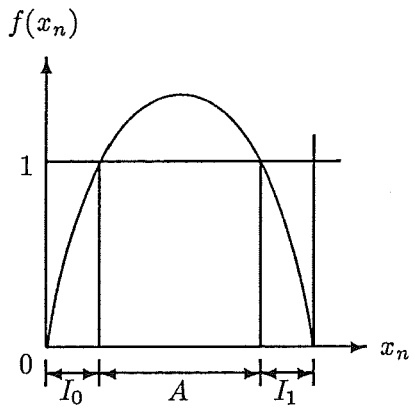


Figure 1.9: Diagrammatic illustration for the fixed point iteration given by $f(x_n) = gx_n(1 - x_n)$ for $g > 2 + \sqrt{5}$ for $x \in [0, 1]$.

After an infinite number of iterations almost all points in $[0, 1]$ leave this interval. However, a set of invariant points remain (i.e. points that do not leave the interval) to form the Cantor set Λ (refer to §2.1). If $x_0 \in \Lambda$ then the entire trajectory of x_0 lies in Λ. The behaviour of such a trajectory can be described exactly using symbolic dynamics.

The itinerary of a point x_0 is defined to be the sequence $S_s(x_0) = \{s_0, s_1, s_2, \cdots\}$ where $s_j = 0$ if $x_j \in I_0$ and $s_j = 1$ if $x_j \in I_1$ (note that $S_s(x_0)$ is a point in the sequence space Σ_2). The itinerary of x_0 is an infinite sequence of the symbols 0 and 1. Hence, S_s is a mapping from Λ to Σ_2 (i.e. $S_s : \Lambda \to \Sigma_2$). Devaney (1987, page 39) shows that S_s is a homomorphism (refer to §2.3), which means that S_s gives an equivalence between the dynamics of f on Λ and σ on Σ_2. Thus

$$S_s \circ f = \sigma \circ S_s \tag{1.51}$$

Since S_s is a homomorphism then f and σ are topologically equivalent, and are therefore completely equivalent in terms of their dynamics (refer to §1.1.2). Thus, the logistic map for $g > 2+\sqrt{5}$ has the properties of the shift map, which in summary are:

- the number of periodic orbits of period n is 2^n.
- f has a dense trajectory in Λ.

Using symbolic dynamics it is possible to completely understand the dynamics of the logistic map for $g > 2 + \sqrt{5}$. For other values of the parameter g the situation is more complicated and a more elaborate version of symbolic dynamics, termed *Kneading theory*, is required (*cf.* Guckenheimer and Holmes 1983; Devaney 1987).

1.1.8 Chaos

Deterministic chaos is a phenomenon related to the occurrence of randomness and unpredictability in completely deterministic systems. This phenomenon has been

termed 'dynamical stochasticity', 'deterministic chaos', 'self-generated noise', 'intrinsic stochasticity' and 'Hamiltonian stochasticity' by various authors (*cf.* Hao 1984). However, the single word 'chaos' has become entrenched in the present day literature as the term used to describe this phenomenon.

There are many possible definitions of chaos, ranging from measure theoretic notions of randomness in ergodic theory (Eckmann and Ruelle 1985) to topological approaches (Collet and Eckmann 1980). Many ODEs exhibit seemingly random or chaotic behaviour. In fact such ODEs form a dense set of the set of all ODEs (Hirsch 1984). Much effort has been expended attempting to categorize and understand this behaviour, but complete success has occurred in only a few cases (Collet and Eckmann 1980). Most chaotic systems are not well understood. This subsection discusses a set of mathematical conditions which have been found useful in defining chaotic behaviour. The main use of this definition at present is to study how systems spawn chaotic dynamics as the system is perturbed from a non-chaotic regime to a chaotic regime. This (hopefully) will provide greater understanding of chaotic behaviour in general. To begin this discussion two mathematical notions are required: sensitive dependence on initial conditions and topological transitivity.

A mapping $f : J \to J$ possesses *sensitive dependence on initial conditions* if there exists $\delta > 0$ such that, for any $x \in J$ and for any neighbourhood N_x of x, there exists $y \in N_x$ and $n \geq 0$ such that $\mid f^n(x) - f^n(y) \mid > \delta$ (Devaney 1987). A map possesses sensitive dependence on initial conditions if there exist points arbitrarily close to x which separate from x by at least δ under iteration of f. Not all points near x need separate from x under iteration, but there must be at least one such point in every neighbourhood of x. If a map possesses sensitive dependence on initial conditions, then for all practical purposes, the dynamics of the map defy numerical computation. Small errors in computation, which are introduced by round-off, may become magnified upon iteration. The results of numerical computation of a trajectory no matter how accurate, may bear no resemblance whatsoever to the real trajectory (*cf.* §6.7).

A mapping $f : J \to J$ is said to be *topologically transitive* if for any pair of open sets $U, V \subset J$ there exists $k > 0$ such that $f^k(U) \cap V \neq \emptyset$ (Collet and Eckmann 1980). A topologically transitive map has points which move under iteration from one arbitrarily small neighbourhood to any other. Consequently, the dynamical system cannot be decomposed into two disjoint sets which are invariant under iteration of the map. If a map possesses a dense trajectory then it is topologically transitive, the converse is also true (Collet and Eckmann 1980).

The following definition for chaos applies to a large number of important examples because in many cases it is possible to verify (Guckenheimer *et al.* 1977; Collet and Eckmann 1980; Devaney 1987). The mapping $f : V \to V$ is said to be chaotic on V if:

- f has sensitive dependence on initial conditions.
- f is topologically transitive.
- Periodic orbits form a dense subset of V.

A chaotic map possesses three features: unpredictability, indecomposability and an element of regularity. A chaotic map is unpredictable because of the sensitive depen-

dence on initial conditions. It cannot be broken down or decomposed into more than one subsystem (i.e. more than one invariant subset) which do not intersect under iteration of f, and there is an element of regularity, periodic orbits form a dense subset. Note that if a system exhibiting sensitive dependence on initial conditions is bounded, then the only trajectories possible behave randomly or stochasticly.

Consider as an example, the shift map σ (defined in §1.1.7), which is chaotic on the sequence space Σ_2. To show that the shift map is chaotic it is sufficient to establish that: σ has sensitive dependence on initial conditions, σ is topologically transitive and σ has periodic trajectories which are dense in Σ_2. In §1.1.7 the shift map is shown to be topologically equivalent to the logistic map for $g > 2 + \sqrt{5}$, therefore the dynamics of the logistic map on the invariant set Λ are also chaotic.

To show that the shift map exhibits sensitive dependence on initial conditions consider two sequences which differ after the i^{th} position. After one iteration of the shift map the two sequences differ after the $(i-1)^{th}$ position (i.e. the two sequences correspond to points in Σ_2 which are further apart). The two points in sequence space that correspond to these sequences separate further apart after each iteration, revealing a sensitive dependence on initial conditions. The shift map is topologically transitive because there exist trajectories which wind densely around sequence space. To see this consider the sequence

$$S_s = (01)(00\ 01\ 10\ 11)(000\ 001\cdots \tag{1.52}$$

where S_s is constructed by successively listing all sequences of 0s and 1s of length n, then of length $n+1$, etc. Some iterate of the shift map applied to S_s can yield any sequence in an arbitrarily large number of places. Therefore this trajectory winds densely around sequence space. The periodic orbits of the shift map form a dense subset of sequence space, since it is possible to find a repeating sequence which is arbitrarily close to any other sequence. This is achieved by choosing the symbols of the repeating sequence to be the same as the beginning symbols of the arbitrary sequence. The repeating sequence can be made arbitrary close by increasing the period of the repeating sequence.

It is difficult to be sure of the presence of chaos in a real system. Observation or simulation over a finite time cannot distinguish between a trajectory of long period and an aperiodic trajectory. For this reason many experimentalists such as Rössler (1976), Ruelle (1980) and Glass *et al.* (1987, page 10) use the term chaos rather loosely to include any situation in which there are trajectories of long periods together with apparently sensitive dependence on initial conditions. Experimentally this is an appropriate definition, since it is only possible to observe a system for a finite time. There is no practical difference between trajectories of long periods and aperiodic trajectories, if both seem to exhibit sensitive dependence on initial conditions.

1.1.9 High Dimensional Dynamics

Flows in a high number of dimensions are generally considerably richer than flows in a low number of dimensions (Hirsch 1984). The approach used to analyse high dimensional flows has been to classify them depending on if they are structurally stable or not (Chua *et al.* 1983, page 698). Peixoto (1962) set criteria for structural stability of flows on two-dimensional manifolds (*cf.* Smale 1967; Mosser 1973). Peixoto's theorem (also known as Peixoto's general density theorem) states that a

flow on a two-dimensional manifold M with Ω as its nonwandering set (a generalisation of fixed points and periodic orbits; refer to §2.1) is structurally stable if and only if the flow satisfies the following conditions:

- The number of fixed points and periodic orbits are finite, and each is hyperbolic.
- There are no trajectories joining fixed points (i.e. no homoclinic or heteroclinic trajectories).
- The nonwandering set Ω consists of fixed points and periodic orbits only.

Moreover, Peixoto (1962) showed that structurally stable flows on two-dimensional manifolds form a dense subset of the set of all flows on two-dimensional manifolds. Smale (1967) and others attempted to extend Peixoto's theorem to higher dimensions (*cf.* Moser 1973; Chillingworth 1976, page 231; Abraham and Marsden 1978). They considered diffeomorphisms which satisfy the Peixoto conditions but with the second condition replaced by a transversality condition (i.e. for any points $p, q \in \Omega$, the stable manifold $W^s(p)$ and the unstable manifold $W^u(q)$ can intersect transversally). Such systems are called *Morse-Smale* (M-S) systems (Chillingworth 1976, page 231). According to Abraham and Marsden (1978, §7.5), Palis and Smale proved M-S systems to be structurally stable. They thought that M-S systems would provide a logical extension to Peixoto's theorem in higher dimensions. However, this was proved wrong when two counter examples were found (i.e. systems that were not M-S but were structurally stable). The first example is by Anosov (Hirsch 1984, page 36), and is called the *hyperbolic toral automorphism* (also known as Anosov diffeomorphism), the second example is by Smale (1967) and is called *Smale's horseshoe* map (Smale's horseshoe map is described in detail in §1.1.9.1). The hyperbolic toral automorphism, Smale's horseshoe map and other examples revealed that a wealth of structurally stable dynamical behaviour can arise in high-dimensional state spaces that cannot exist in low dimensional state spaces.

Smale (1966) showed that structurally stable flows do not form a dense subset of the set of all flows (*cf.* Chillingworth 1976, page 244). However, while structurally stable flows are not dense, Smale considered that they might be able to be characterised by a convenient set of properties. According to Hirsch (1984 page 39), Smale put forward a list of such properties which is revised as counter-examples are found. Presently the gap between known necessary properties and known sufficient properties is still unbridged. One property which is known to be sufficient (put forward by Smale 1967) is called *axiom A* (*cf.* Bowen 1980). A diffeomorphism f is said to satisfy axiom A if the nonwandering set Ω of f is hyperbolic and the periodic orbits of f form a dense subset of Ω (Smale 1967). A flow satisfying axiom A is not necessary an attractor: Smale's horseshoe satisfies axiom A but is not an attractor. According to Holmes and Marsden (1982), while horseshoes are not strange attractors, they are often visible and behave like them in numerical experiments (perhaps due to noise). The solenoid attractor (Devaney 1987, page 198) and the Plykin attractor (Guckenheimer and Holmes 1983, page 264) are examples of axiom A attractors. According to Zeeman (1983, page 42) non-axiom A attractors are relatively unexplored. It appears that most flows which have been derived from physical phenomena and have strange attractors are non-axiom A. Two of the most studied examples are the Hénon (1976) attractor and the Lorenz (1963) attractor (*cf.* Sparrow 1982).

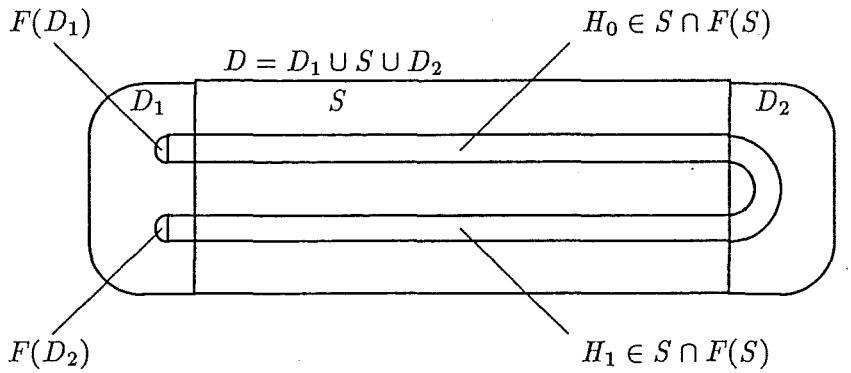


Figure 1.10: Smale's horseshoe map. The region D consists of three components: a central square S of unity side length, and two semicircles D_1 and D_2 at either end. The horseshoe map, maps the region D into itself.

1.1.9.1 Smale's Horseshoe Map

Smale's horseshoe map, denoted F, can be explained by examining the stadium shaped region D shown in figure 1.10. The region D consists of three components: a central square S of unity side length and two semicircles D_1 and D_2 at either end. The horseshoe map maps the region D into itself. It is convenient to consider the horseshoe map in two steps. First, the region S is linearly contracted in the vertical direction by a factor $\delta < 1/2$, and expanded in the horizontal direction by a factor $1/\delta$ so that S becomes long and thin. Then S is folded and placed back inside D as shown in figure 1.10. Second, the semicircular regions D_1 and D_2 are contracted and mapped inside D_1 as shown in figure 1.10. Since F contracts the region D_1 into itself, F has a stable fixed point p in D_1 (i.e. $F^n(q) \to p$ as $n \to \infty$ for all $q \in D_1$), and since $F(D_2) \subset D_1$, all trajectories of points in D_2 behaviour likewise. Similarly, if $q \in S$ but $F^k \notin S$ for some $k > 0$, then $F^k(q) \to p$ as $n \to \infty$. Consequently to understand the trajectories of F, it suffices to consider the set of points, denoted Λ, whose trajectories lie for all time in S (i.e. the invariant set Λ)

$$\Lambda = \{q \in S : F^k(q) \in S \text{ for all } k \in \mathbf{Z}\} \tag{1.53}$$

where $\mathbf{Z}$ is the set of integers (refer to §2.1). It is convenient to make Λ the intersection of the two sets Λ_+ and Λ_-, where

$$\begin{aligned} \Lambda_+ &= \{q : F^k(q) \in S \text{ for } k = 0, 1, 2, \cdots\} \\ \Lambda_- &= \{q : F^{-k}(q) \in S \text{ for } k = 1, 2, \cdots\} \end{aligned} \tag{1.54}$$

The region $F(S) \cap S$ consists of two horizontal rectangles, denoted H_0 and H_1, of height δ and of unit length. The region $F^{-1}(F(S) \cap S)$ (i.e. the preimage of the two horizontal rectangles H_0 and H_1; refer to §2.3) consists of two vertical rectangles, denoted V_0 and V_1, of width δ and of unit height (refer to figure 1.11), the region $F^{-2}(F(S) \cap S)$ consists of four vertical rectangles, denoted V_{00}, V_{01}, V_{10} and V_{11}, of width δ^2 (refer to figure 1.11), the region $F^{-3}(F(S) \cap S)$ consists of eight rectangles, etc. If $F(q)$ lies in S, then $q \in V_0 \cup V_1$ as all other points in S are mapped out of S and into $D_1 \cup D_2$, if $F^2(q)$ lies in S, then $q \in V_{00} \cup V_{01} \cup V_{10} \cup V_{11}$, etc.

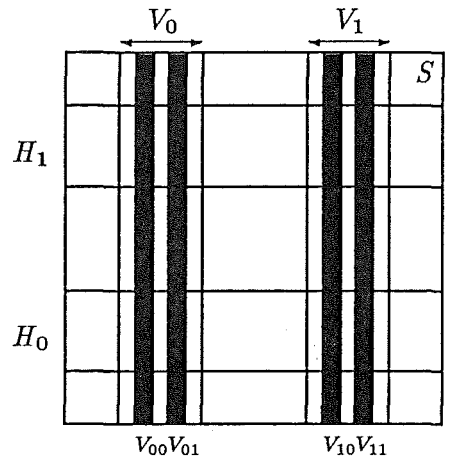


Figure 1.11: The region $F(S) \cap S$ consists of two horizontal rectangles, denoted H_0 and H_1, of height δ and unity length. The region $F^{-1}(F(S) \cap S)$ consists of two vertical rectangles, denoted V_0 and V_1, of width δ and unity height, the region $F^{-2}(F(S) \cap S)$ consists of four vertical rectangles, denoted V_{00}, V_{01}, V_{10} and V_{11}, of width δ^2, etc.

By inductive reasoning, if $V \in V_0 \cup V_1$ is any vertical rectangle of width w connecting the upper and lower boundaries of S, then $F^{-1}(V)$ is a pair of vertical rectangles of width δw, one in V_0 and one in V_1. Consequently, $F^{-2}(V)$ consists of four vertical rectangles, each of width $\delta^2 w$, $F^{-3}(V)$ consists of eight vertical rectangles of width $\delta^3 w$, etc. (note the similarity with the construction of the Cantor set described in §2.1). Using an argument similar to that developed in §1.1.7 it is possible to show that the set Λ_+ is the Cartesian product of a Cantor set and a vertical line segment of unit length (Guckenheimer and Holmes 1983, §5.1). Arguing entirely analogously, it is possible to show that Λ_- is the Cartesian product of a Cantor set and a horizontal line of unit length. The entire invariant set Λ is the intersection of Λ_+ and Λ_-.

Symbolic dynamics can be invoked to describe the dynamics of the horseshoe map F on Λ. First, choose any vertical line segment of length $\ell \leq 1$ in Λ_+. Note that, $F^k(\ell)$ is a vertical line segment of length $\delta^k \ell$ in either V_0 or V_1. It is possible to attach an infinite sequence $\{s_0, s_1, \cdots\}$ of 0's or 1's to any point on ℓ according to the rule $s_j = 0$ if $F^j(\ell) \subset V_0$ and $s_j = 1$ if $F^j(\ell) \subset V_1$ for $j = 0, 1, \cdots$ (Guckenheimer and Holmes 1983, §5.1; Devaney 1987). The number s_0 tells in which vertical rectangle (i.e. V_0 or V_1) the line ℓ is located, s_1 tells which vertical rectangle its image is located, etc. Second, choose any horizontal line segment of length $h \leq 1$ in Λ_-. Attach a sequence of integers, for convenience write this sequence as $\{\cdots, s_{-2}, s_{-1}\}$, to any point on h according to the rule $s_{-j} = 0$ if $F^{-j}(h) \subset V_0$ and $s_{-j} = 1$ if $F^{-j}(h) \subset V_1$ for $j = 1, 2, \cdots$. Consequently, if q is any point on Λ, it can be associated with a pair of sequences. One sequence gives the itinerary of the forward trajectory of q, the other gives the backward trajectory. Both sequences can be amalgamated into one doubly-infinite sequence of 0's and 1's (Guckenheimer and Holmes 1983, §5.1; Devaney 1987). That is, the itinerary $S_s(q) = \{\cdots, s_{-2}, s_{-1}, s_0, s_1, \cdots\}$, defined by the rule $s_j = 0$ if $F^j(q) \subset V_0$ and $s_j = 1$ if $F^j(q) \subset V_1$ for $j = \cdots, -1, 0, 1, \cdots$. This then gives the symbolic dynamics on Λ. Let Σ_2 denote the set of all doubly-infinite

sequences of 0's and 1's. Define the double-sided shift map σ by

$$\sigma(\cdots, s_{-2}, s_{-1}, s_0, s_1, \cdots) = (\cdots, s_{-1}, s_0, s_1, s_2, \cdots) \tag{1.55}$$

where σ moves each sequence in Σ_2 one unit to the left. Devaney (1987, page 178) proves there exists a topological conjugacy between F on Λ and σ on Σ_2. All properties that hold for the (one-sided) shift map discussed in §1.1.7 also holds for the double-sided shift map σ (1.55), namely

- σ has sensitive dependence on initial conditions.
- σ is topologically transitive.
- periodic points are dense in Λ.

Thus, Smale's horseshoe map F is chaotic in the sense discussed in §1.1.8 on the invariant set Λ (*cf.* Smale 1967; Chua *et al.* 1983b; Hirsch 1984; Kloeden and Mees 1985; Chua 1987).

1.1.10 Homoclinic Trajectories

The significance of the concepts raised in the previous subsections can be made clearer through an example. In this subsection the pendulum example discussed at the end of §1.1.2 is revisited.

The last two paragraphs of §1.1.2 discussed the effect of adding friction to the frictionless pendulum. It was found that the stable and unstable invariant manifolds which are coincident for a frictionless pendulum separate when friction is added. If instead of adding friction, one perturbs the length of the pendulum periodically, the stable and unstable invariant manifolds again separate, but in a significantly different way, as is outlined in the following discussion. For the example discussed here the length of the pendulum ℓ is given by the function $\ell = L + \varepsilon \sin wt$, where L, w and ε are constants. The angular acceleration of the pendulum is

$$\ddot{\theta} = \frac{-g}{(L + \varepsilon \sin(wt))} \sin\theta \tag{1.56}$$

and provided ε is small, the angular acceleration can be adequately approximated by the first two terms of its binomial expansion (*cf.* Chua 1987), i.e.

$$\ddot{\theta} = \left(\frac{-g}{L} + \frac{-\varepsilon g}{L^2} \sin wt\right) \sin\theta \tag{1.57}$$

If g/L is set to unity and $\varepsilon g/L^2 = \epsilon$, the angular acceleration simplifies to

$$\ddot{\theta} = -\sin\theta - \epsilon \sin(wt) \sin\theta \tag{1.58}$$

For $\epsilon = 0$ (1.58) characterises the unperturbed frictionless pendulum (1.4) while for $\epsilon > 0$ (1.58) is a non-autonomous dynamical system. As already explained in §1.1.4 a non-autonomous system can always be transformed into an autonomous system by adding an extra equation. Let $x_1 = \theta$, $x_2 = \dot{x}_1$, $x_3 = t$, then the perturbed pendulum characterised by (1.58) can be written as a system of three first order equations

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= -\sin\theta - \varepsilon \sin(wx_3) \sin\theta \\ \dot{x}_3 &= 1 \end{aligned} \tag{1.59}$$

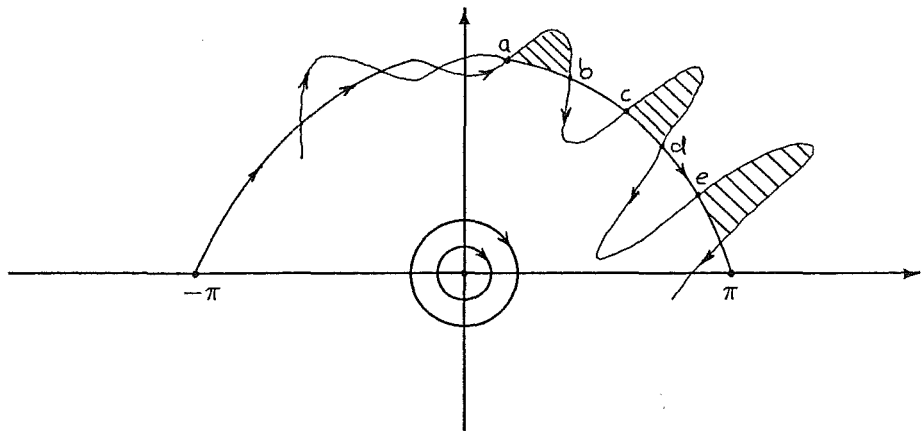


Figure 1.12: Poincaré map obtained when the length of the pendulum is perturbed by a sinusoidal time function. The stable and unstable invariant manifolds of the hyperbolic fixed points twist around each other (wildly) and intersect transversally (i.e. do not intersect tangentially).

The state space characterising (1.59) is three-dimensional, with coordinates x_1, x_2 and x_3. By considering the intersection of the flow with a plane (denoted by Σ), a two-dimensional cross section of state space is formed. Let the orientation of Σ be such that it intersects both the x_1 and x_2 coordinate axes. Σ divides the three-dimensional state space into two regions Σ^+ and Σ^-. If points are plotted on Σ where the flow intersects Σ only when the flow is traversing from say region Σ^+ to Σ^- (or Σ^- to Σ^+) then Σ forms a Poincaré map (refer to §1.1.4). Figure 1.12 shows a plot of the Poincaré map for ϵ equal to some small positive real number.

Fixed points of the Poincaré map represent periodic orbits of the three-dimensional flow. Fixed points occur along the x_1 axis in figure 1.12 separated by a distance π and are alternatively hyperbolic and elliptic (which is similar to the frictionless pendulum discussed in §1.1.2). The stable and unstable invariant manifolds emanating from the hyperbolic periodic orbits intersect the Poincaré map forming curves. Of significance is that the stable and unstable invariant manifolds twist about each other, and actually intersect each other transversally (i.e. do not intersect tangentially) at an infinite number of points on the Poincaré map (*cf.* Dragt and Finn 1976; Guckenheimer and Holmes 1983; Moon 1987). These intersection points are called transverse heteroclinic points (e.g. the points marked 'a', 'b', 'c' and 'd' in figure 1.12 are points of transverse intersection). The flow takes point 'a' in figure 1.12 to point 'b', then to point 'c', etc. Since a point on an invariant manifold remains on the invariant manifold for all time, a point of intersection of the stable and unstable invariant manifolds belongs to both manifolds, and thus must remain on both manifolds for all time. The points in between the stable and unstable invariant manifolds, represented by the shaded regions in figure 1.12, remain trapped between the invariant manifolds. The shaded region between points 'a' and 'b' in figure 1.12 moves under the flow to the shaded region between the points 'b' and 'c', then to the shaded region between points 'c' and 'd', etc. (*cf.* Dragt and Finn 1976).

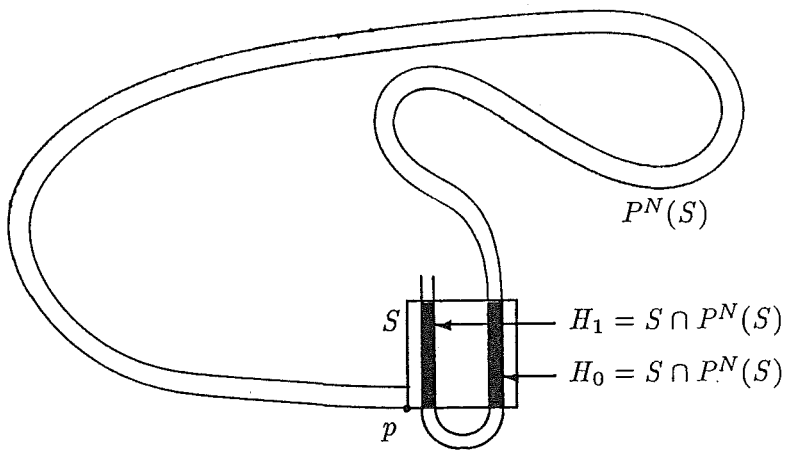


Figure 1.13: Close up view in the neighbourhood of the homoclinic point. Smale (1967) proved that what is happening near this homoclinic point is not very different from what is happening on and near trajectories threading through Smale's horseshoe map nonwandering set.

Since the velocity function (1.58) is time-periodic, the transformed system (1.59) is periodic in x_3 with period 2π. It is therefore possible to transform (1.59) from $\mathbf{R}^3$ Euclidean state space to cylindrical state space $\mathbf{R}^2 \times \mathbf{S}$ (refer to §1.1.4). A Poincaré map of the transformed system is

$$\Sigma = \{(x_1, x_2, x_3), x_1, x_2 \in \mathbf{R}, x_3 \in \mathbf{S} : x_3 = \theta\} \tag{1.60}$$

where $\theta \in \mathbf{S}$ is an arbitrary constant. The infinite set of heteroclinic points in figure 1.12 are transformed into a single homoclinic point on a cylindrical state space. Figure 1.13 depicts an expanded view in the neighbourhood of the homoclinic point p for the pendulum. A small square S positioned near the homoclinic point p becomes more and more stretched after each iteration of the Poincaré map P (note that P is a diffeomorphism). After N iterations the region $P^N(S)$ intersects S and because of the way the invariant manifolds are twisted about each other, this intersection is characterised by a number of horizontal rectangles H_i. The preimage (refer to §2.3) of each H_i is a vertical rectangle V_i (i.e. $V_i = P^{-N}(H_i)$ where $H_i \in S$, and $V_i \in S$), reminiscent of Smales horseshoe map. Smale (1967) proved that what is happening near such homoclinic points is not very different from what is happening on and near trajectories threading through the horseshoe map's invariant set. He showed that in the neighbourhood of the homoclinic point p of the diffeomorphism P, there is a Cantor set Λ with $p \in \Lambda$ and an integer $N > 0$ such that $P^N(\Lambda) = \Lambda$, and P^N restricted to Λ is topologically equivalent to the shift map discussed in §1.1.9.1. In other words, the N^{th} iterate of the diffeomorphism acts the same as Smale's horseshoe map. Direct corollaries are that in an arbitrarily small neighbourhood of the homoclinic trajectory there are periodic orbits, and that trajectories near the homoclinic trajectory exhibit sensitive dependence on initial conditions (*cf.* Smale 1967; Mosser 1973; Mees 1981; Chua *et al.* 1983b; Guchenheimer and Holmes 1983; Moon 1987).

Since the invariant set associated with the homoclinic point p is a Cantor set, the invariant set of the flow near a homoclinic trajectory is topologically the Cartesian

product of a Cantor set and a line. It is therefore a complicated sort of surface with a non-integer dimension (Ruelle 1980). It is termed a *strange invariant set*, and if this set is an attractor it is a strange attractor (refer to §1.1.3). Smale's horseshoe has a nonattracting strange invariant set. Nonattracting strange invariant sets are not directly observable, but they typically give rise to very complicated transient behaviour (Holmes and Marsden 1982). Trajectories wander around in the tangle about the homoclinic trajectory until eventually they get pulled out of the homoclinic region and into some other, perhaps regular, attractor. Such strange attractors, are at least, one way of generating chaos. This is why there is much interest in heteroclinic and homoclinic trajectories (*cf.* Nicolis 1986a).

1.1.11 Bifurcation Scenarios Preceding Chaos

Suppose the behaviour of a dynamical system depends on an external controllable parameter, and that for some value of the parameter the system exhibits regular dynamical behaviour. Suppose the parameter is initially set to a value such that the dynamical system exhibits regular behaviour. As the parameter is changed from this value the qualitative behaviour of the system may change too (Eckmann 1981, page 643). After a finite or infinite succession of such changes or bifurcations the system may exhibit chaotic behaviour. According to Eckmann (1981) this sequence of bifurcations, ultimately culminating in chaos, is called a *scenario preceding chaos.* There are three widely known scenarios documented in the literature (Eckmann 1981; Ott 1981; McCauley 1988): the Ruelle and Takens (1971) scenario, the Feigenbaum (1978) scenario and the Pomeau and Manneville (1980) scenario.

The Ruelle and Takens (1971) scenario consists of two or three bifurcations. As explained in §1.1.2 one of the attractors that become possible in a state space of three or more dimensions is quasi-periodic motion on a torus. Such tori can form through a bifurcation from a periodic orbit. Ruelle and Takens (1971) showed that a two-dimensional torus is structurally stable, but this is not so for tori of three or more dimensions. These are fragile objects disappearing upon the action of a small perturbation; their destruction may give rise to a strange attractor. Such bifurcations preceding chaos have been observed in models of physical systems and in experiments, notably in fluid dynamics (Swinney and Gollub 1978).

The Feigenbaum (1978) scenario consists of an infinite sequence of bifurcations, each bifurcation occurring at a well defined parameter value g (refer to §1.1.6). Each bifurcation leads to longer periodic orbits, whose period doubles at each consecutive bifurcation. The bifurcations accumulate at a particular parameter value g_∞, after which one obtains chaotic behaviour. There exists extensive experimental evidence that large classes of physical systems evolve to chaos through this scenario (Eckmann 1981; Ott 1981; Hao 1984; Pei *et al.* 1986).

The Pomeau and Manneville (1980) scenario arises through intermittency. The trajectory of a dynamical system becomes increasingly intermittent as a parameter is changed, ultimately leading to chaotic behaviour. It arises when regular behaviour becomes unstable in a particular way. The resulting trajectory has intervals of regular behaviour followed by turbulent bursts. The interval between and the duration of the turbulent bursts are seemingly unpredictable. This scenario has been observed in actual experiments (Ott 1981).

The significance of these scenarios is the existence of universal trends in the approach to the deterministic chaos (Feigenbaum 1978) as a parameter value is changed. According to Nicolis (1986 page 902), "there is no doubt that it constitutes one of the major scientific breakthroughs of the last decade".

1.1.12 Measuring Chaos

A number of quantitative measures have been formulated to characterise chaotic behaviour (*cf.* Kloeden and Mees 1985, §4). This subsection outlines some of the quantitative measures that have been found useful in practice. This subsection discusses Lyapunov exponents, the spatial dimension of an attractor and the (Kolmogorov) entropy of a trajectory. The Lyapunov exponent measures the average rate of expansion or contraction of state space in particular directions. It gives an indication of how sensitive a flow is to initial conditions (Pesin 1977; Schuster 1983). The dimension measure determines the spatial dimension of an attractor in state space (Farmer *et al.* 1983; Grassberger and Procaccia 1984). Many measures of spatial dimension have been proposed, the main differences lie in the way the flow (if taken into account at all) on the attractor is taken into account. A non-integer dimension of an attractor is an indication of a strange attractor. Motion on a strange attractor is likely to look more unpredictable if the dimension of the attractor is high rather than low (Kloeden and Mees 1985, page 75). The (Kolmogorov) entropy measures the rate at which information is generated by the system, or stated differently, entropy measures the rate at which information known about a system (e.g. the accuracy of an initial condition) is lost.

Lyapunov exponents characterise the stability of both regular and chaotic behaviour and are a generalisation of the eigenvalues of a Jacobian matrix (refer to §1.1.2) calculated at points on the flow. Let $m_i(t), i = 1, \cdots, n$ be the eigenvalues of the flow $\phi_t(\mathbf{u})$. The Lyapunov exponents λ_i, $i = 1, \cdots, n$ are defined to be

$$\lambda_i \stackrel{def}{=} \lim_{t \to \infty} \frac{1}{t} \ln \mid m_i(t) \mid \tag{1.61}$$

To gain familiarity with Lyapunov exponents, consider the Lyapunov exponents at a hyperbolic fixed point $\bar{x}$. The flow at the fixed point is given by

$$\phi_t(\bar{x}) = e^{f'(\bar{x})t} \tag{1.62}$$

where $f'(\cdot)$ is the Jacobian matrix of the velocity function f (refer to §1.1.2). It follows that $m_i(t) = e^{\hat{\lambda}_i t}$ and

$$\begin{aligned} \lambda_i &= \lim_{t \to \infty} \frac{1}{t} \ln \mid e^{\hat{\lambda}_i t} \mid \\ &= \mathrm{Re}(\hat{\lambda}_i) \end{aligned} \tag{1.63}$$

where $\mathrm{Re}(a)$ denotes the real part of a. In this special case the Lyapunov exponents are equal to the real part of the eigenvalues of the fixed point (*cf.* Chua *et al.* 1987, page 994). Note that, it is common for λ to denote both the value of eigenvalues and Lyapunov exponents. In this thesis $\hat{\lambda}$ denotes eigenvalues and λ denotes Lyapunov exponents.

The Lyapunov exponent gives the average rate of contraction ($\lambda_i < 0$) or expansion ($\lambda_i > 0$) along particular directions in state space. What is meant by particular

direction is that if the exponents are ordered $\lambda_1 \geq \cdots \geq \lambda_n$, then there exist n linear subspaces (refer to §2.2), $w_1 \supset w_2 \supset \cdots \supset w_n$, with dimensions $w_1 = n$, $w_2 = n-1, \cdots, w_n = 1$, such that almost all (a set of measure zero does not) perturbations in w_i expand or contract (on average) as $e^{\lambda_i t}$. Since a strange attractor possesses sensitive dependence on initial conditions, at least one Lyapunov exponent must be positive. If more than one Lyapunov exponent is positive the flow is termed hyperchaotic (Rössler 1979; Matsumoto *et al.* 1986a).

An attractor is n-dimensional if at every point on the attractor there exists a neighbourhood that looks like an open subset of $\mathbf{R}^n$ (Farmer *et al.* 1983). For example, a periodic orbit is one-dimensional since there exists a neighbourhood of every point on the periodic orbit that looks like a line. The neighbourhood of any point of a strange attractor usually has a fine structure (e.g. the Cartesian product of a Cantor set and a line) and does not resemble Euclidean space (Ruelle 1980). Most known strange attractors have non-integer dimensions. There are several ways to generalize dimension to the non-integer case (Grassberger and Procaccia 1983). Four generalisations are considered in this subsection, they are termed fractal dimension (also termed capacity), information dimension, correlation dimension and Lyapunov dimension.

Fractal dimension (Mandelbrot 1977) which is denoted D_{cap}, is defined by covering an attractor Λ with volume elements (e.g. spheres, cubes, etc.) each with diameter ϵ. Let $N(\epsilon)$ be the number of volume elements needed to cover Λ. As ϵ is made smaller, the sum of the volume elements approaches the volume of Λ. If Λ is d-dimensional then for ϵ small the number of volume elements needed to cover Λ is inversely proportional to ϵ^{-d}, that is, $N(\epsilon) = k\epsilon^{-d}$ for some constant k (cf Chua *et al.* 1987, page 996). The definition of D_{cap} is

$$D_{cap} \stackrel{def}{=} \lim_{\epsilon \to 0} \frac{\ln N(\epsilon)}{\ln(1/\epsilon)} \tag{1.64}$$

D_{cap} is a purely metric concept and does not utilize information about the time behaviour of the dynamical system. Information dimension D_i is a probabilistic dimension measure, defined in terms of the probability of visitation of a trajectory to volume elements covering the attractor. A covering of $N(\epsilon)$ volume elements each with diameter ϵ is made. The definition of D_i is

$$D_i \stackrel{def}{=} \lim_{\epsilon \to 0} \frac{\ln S(\epsilon)}{\ln(1/\epsilon)} \tag{1.65}$$

where

$$S(\epsilon) \stackrel{def}{=} -\sum_{i=1}^{N(\epsilon)} P_i \ln P_i \tag{1.66}$$

and P_i is the probability of the system being in a state contained in the i^{th} volume element of the covering. $S(\epsilon)$ is termed entropy and is the amount of information needed to specify the state of the system to an accuracy of ϵ if the state of the system is known to be on the attractor (*cf.* Pierce 1980). Another probabilistic dimension is the correlation dimension D_c (Chua 1987, page 996), which is defined as

$$D_c \stackrel{def}{=} \lim_{\epsilon \to 0} \frac{\ln C(\epsilon)}{\ln \epsilon} \tag{1.67}$$

where

$$C(\epsilon) \stackrel{def}{=} \lim_{N\to\infty} \frac{1}{n^2}\{\text{number of point pairs } x_i, x_j \text{ such that } \mid x_i - x_j \mid < \epsilon\} \quad (1.68)$$

is the correlation between a set of n points on a particular trajectory. Kaplan and Yorke (1979) conjectured that there is a relationship between the dimension and the Lyapunov exponents of an attractor. This relationship is termed the Lyapunov dimension D_ℓ. To define the Lyapunov dimension order the Lyapunov exponents form the largest to smallest (i.e. $\lambda_1 \geq \cdots \geq \lambda_n$) and let j be the largest integer such that $\lambda_1 + \cdots + \lambda_j \geq 0$. The definition of D_ℓ is (Frederickson *et al.* 1983)

$$D_\ell \stackrel{def}{=} j + \frac{\lambda_1 + \cdots \lambda_j}{\mid \lambda_{j+1} \mid} \quad (1.69)$$

The Lyapunov dimension provides a computationally inexpensive method for calculating the dimension of an attractor (once the Lyapunov exponents are known). Kaplan and York (1979) originally conjectured that $D_\ell = D_{cap}$, and various plausibility arguments were put forward that suggested this. Computer simulations seem to indicate that D_ℓ lies closer to D_i (Lichtenberg and Lieberman 1982, page 395; Grassberger and Procaccia 1983). It is amazing that seemingly static properties like dimension can be related to dynamical properties like Lyapunov exponents.

To define the Kolmogorov entropy consider the trajectory $\mathbf{x}(t, \mathbf{u})$ of a dynamical system on a strange attractor. Suppose that the d-dimensional state space is partitioned into boxes of size ϵ^d and the state of the system is measured at intervals of time τ. Let $P_{i_0,\cdots,i_n}$ be the joint probability that $\mathbf{x}(t_0, \mathbf{u})$ is in box i_0, that $\mathbf{x}(t_0+\tau, \mathbf{u})$ is in box i_1, etc., and that $\mathbf{x}(t_0 + n\tau, \mathbf{u})$ is in box i_n. Then

$$K_n = -\sum_{i_0,\cdots,i_n} P_{i_0,\cdots,i_n} \ln P_{i_0,\cdots,i_n} \quad (1.70)$$

is the information needed to locate the system on any particular trajectory, $i_0, \cdots, i_n$ with precision ϵ. The Kolmogorov entropy is $K_{n+1} - K_n$, which is the additional information needed to predict in which cell i_{n+1} the system will be if it is known that it was previously in $i_0, \cdots, i_n$. Thus, the Kolmogorov entropy measures the loss of information known about a system as it evolves from time n to time $n + 1$ (*cf.* Schuster 1983; Eckmann and Rulle 1985, page 637).

1.2 Conservative Dynamical Systems

A conservative dynamical system is a dynamical system where a special constraint function, termed the *Hamiltonian function*, is conserved (i.e. equals a constant) for all time. The Hamiltonian function is based on a generalised principle. It provides a method for bringing a number of problems from different disciplines into a general framework (Percival and Richards 1982). §1.2.1 describes this generalised approach for classical mechanics. In classical mechanics the Hamiltonian function is usually equal to the total energy of the system (Percival and Richards 1982). Hence a conservative dynamical system is often considered to be a system which conserves energy (i.e. has no energy loss). This is equivalent to saying that the volume of any region in state space is conserved, as the region evolves according to the system equations

(i.e. there are no sinks or sources) (Marion 1970, page 229). Conservative systems are special as they lie in between systems that loose energy and have attractors (i.e. dissipative systems), and systems that gain energy and ultimately have unbounded motion. This has important consequences for the stability and robustness of such systems to small perturbations of the trajectories and of the system equations. If an integrable (i.e. solvable analytically) conservative system is perturbed slightly by adding a nonlinear term (but the entire system remains a conservative system) the behaviour of the system becomes in general complicated. For example, the three body problem (Marion 1970) can be formulated as an integrable conservative system perturbed by a nonlinear term. This makes the three body problem essentially unsolvable (except for certain restricted cases). The nature of this complicated behaviour resulting from a small nonlinear perturbation is discussed in §1.2.2.

1.2.1 Hamilton's Principle in Classical Mechanics

Hamilton's principle is the notion that a single particle whose total energy is conserved always traces the path $\mathbf{x}(t)$ that makes the action integral

$$S = \int_{t_1}^{t_2} L(\mathbf{x}(t), \mathbf{v}(t))dt \tag{1.71}$$

have a stationary value (i.e. the path $\mathbf{x}(t)$ that makes the derivative of S zero), where $\mathbf{v}(t)$ is the velocity of the particle, $\mathbf{x}(t)$ is the displacement or path of the particle, t_1 and t_2 are the start and finish times and

$$L(\mathbf{x}(t), \mathbf{v}(t)) \stackrel{def}{=} T(\mathbf{v}(t)) - V(\mathbf{x}(t)) \tag{1.72}$$

is termed the Lagrangian of the particle (*cf.* Arnold 1978), where $T(\mathbf{v}(t))$ is the total kinetic energy of the system and $V(\mathbf{v}(t))$ is the total potential energy of the system (Marion 1970, page 197). Invoking calculus of variations (*cf.* Arnold 1978) it is possible to show that for S to be stationary

$$\frac{d}{dt}\left(\frac{\delta L(\mathbf{x}(t), \mathbf{v}(t))}{\delta \mathbf{v}(t)}\right) - \frac{\delta L(\mathbf{x}(t), \mathbf{v}(t))}{\delta \mathbf{x}(t)} = 0 \tag{1.73}$$

and for S to be a minimum the mass of the particle must be positive. Equation (1.73) is the general condition that Hamilton's Principle imposes on $\mathbf{x}(t)$. If the motion of a particle can be described by a Lagrangian then the quantity

$$\mathbf{p}(t)\mathbf{v}(t) - L(\mathbf{x}(t), \mathbf{v}(t)) \tag{1.74}$$

where $\mathbf{p}(t)$ is the momentum of the particle, is a constant of the motion and is equal to the total energy of the particle (Marion 1970). A partial Legendre transform (*cf.* Arnold 1978) with respect to the velocity in (1.74) defines the Hamiltonian function $H(\mathbf{x}(t), \mathbf{p}(t))$

$$H(\mathbf{x}(t), \mathbf{p}(t)) \stackrel{def}{=} \frac{\delta L(\mathbf{x}(t), \mathbf{v}(t))}{\delta \mathbf{v}(t)} - L(\mathbf{x}(t), \mathbf{v}(t)) \tag{1.75}$$

or expressed in the standard or canonical form is

$$\frac{\delta H(\mathbf{x}(t), \mathbf{p}(t))}{\delta \mathbf{p}(t)} = \mathbf{v}(t)$$

$$\begin{aligned} -\frac{\delta H(\mathbf{x}(t),\mathbf{p}(t))}{\delta \mathbf{x}} &= \frac{d\mathbf{p}(t)}{dt} \\ \frac{dH(\mathbf{x}(t),\mathbf{p}(t))}{dt} &= -\frac{dL(\mathbf{x}(t),\mathbf{v}(t))}{dt} \end{aligned} \tag{1.76}$$

These equations are termed *Hamilton's equations* or the *Hamiltonian* (Marion 1970, page 220). Because of the constraint imposed by the Hamiltonian, a conservative system of order $2n$ (i.e. requiring a state space of dimension $2n$) has only n degrees of freedom (refer to §1.1). Besides classical mechanical systems there are electrical, biological, meteorological and economic conservative systems (Percival and Richards 1982). More generally the variable $\mathbf{x}$ is usually replaced by $\mathbf{q}$, and is termed the *generalised coordinate* and the quantity $\mathbf{p}$ is termed the *conjugate momentum.* The pair $(\mathbf{q},\mathbf{p})$ are termed conjugate variables (Marion 1970, page 200). In general $\mathbf{q}$ need not represent configuration (distance in mechanics) nor need $\mathbf{p}$ be a physical momentum, although they frequently have this meaning. Note that, if the value of the Hamiltonian is dependent on time then the system it characterises is not conservative.

1.2.2 KAM Theory

In this subsection KAM (Kolmogorov, Arnold, Moser) theory is intuitively described (Lichtenberg and Lieberman 1983, Chapter 3). It is usually found that the motion described by a Hamiltonian is easier to solve if it is first transformed into *action-angle variables* (action-angle variables are explained below) (Percival and Richards 1982). Consider a Hamiltonian characterising two coupled oscillators transformed into action-angle variables (J,φ)

$$H(J_1,J_2,\varphi_1,\varphi_2) = H_0(J_1,J_2,\varphi_1,\varphi_2) \tag{1.77}$$

The trajectories in state space are characterised by (1.77) when $H(J_1,J_2,\varphi_1,\varphi_2)$ is set to some real constant. The trajectories lie on a torus where (φ_1,φ_2) are the angle, and (J_1,J_2) are the radii specifying the torus (refer to figure 1.14). The motion of a trajectory on the torus consists of the Cartesian product of two periodic orbits. The period p (or frequency $f = 1/p$) of each periodic orbit p_1,p_2 is related to each other through equation (1.77). The period of and the relationship (i.e. the ratio p_1/p_2) between the periods of the periodic orbits are determined by the value of $H(J_1,J_2,\varphi_1,\varphi_2)$. Each value of $H(J_1,J_2,\varphi_1,\varphi_2)$ corresponds to a different size torus in state space. It is useful to think of this state space as consisting of an infinite number of tori, each fitting inside the other, and each corresponding to a particular value of $H(J_1,J_2,\varphi_1,\varphi_2)$. Now consider a small nonlinear perturbation to $H(J_1,J_2,\varphi_1,\varphi_2)$, i.e.

$$H(J_1,J_2,\varphi_1,\varphi_2) = H_0(J_1,J_2,\varphi_1,\varphi_2) + V(J_1,J_2,\varphi_1,\varphi_2) \tag{1.78}$$

The nonlinear term $V(J_1,J_2,\varphi_1,\varphi_2)$ either perturbs the shape of or destroys the tori on which the trajectories flow. KAM theory proves that most of the tori of the unperturbed system, where the ratio between the periods of the periodic orbits are incommensurate (i.e. p_1/p_2 is an irrational number) continue to exist, being only slightly distorted by the perturbation. These tori are known as KAM curves (Lichtenberg and Lieberman 1983, Chapter 3). On the other hand the tori bearing

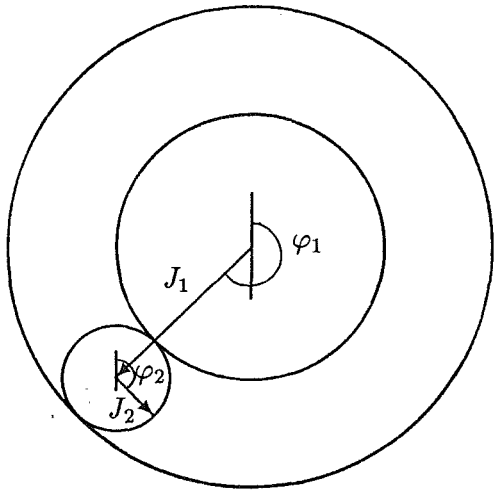


Figure 1.14: Trajectories characterised by (1.77) lie on a torus, where (φ_1, φ_2) are the angle, and (J_1, J_2) are the radii specifying the torus.

trajectory motion with commensurate periods or with incommensurate periods whose ratio is approximated well by (r/s) (where r and s are relatively small integers), are grossly deformed by the perturbation and no longer remain close to the tori of the unperturbed system (*cf.* Arnold 1963; Walker and Ford 1969).

A Poincaré map of (1.77) and (1.78) can be constructed in the way described in §1.1.4 for a non-autonomous system. A Poincaré map is specified by the intersection of trajectories with the surface $\Sigma = \{(J_1, J_2, \varphi_1, \varphi_2), J_1, J_2 \in \mathbf{R}, \varphi_1, \varphi_2 \in \mathrm{S} : \varphi_1 = \theta\}$, where θ is arbitrary. For tori where the ratio between the periods of the periodic orbits are commensurate, the trajectories on the tori cut the surface of section forming the Poincaré map at a finite number (i.e. p_1/p_2 points, where $p_1 < p_2$) of points only. A small nonlinear perturbation of the Hamiltonian changes these points into an alternating series of elliptic and hyperbolic fixed points, with trajectories encircling the elliptic fixed points and a separatrix trajectory connecting the hyperbolic points. This phenomenon is known as a *primary resonance* (Lichtenberg and Lieberman 1983, Chapter 3). A detailed examination in the neighbourhood of any elliptic fixed point reveals a higher order resonance, which have their own motion similar to that described for the primary resonance but on a finer scale. Seemingly stochastic regions form around the neighbourhood of hyperbolic fixed points, and these stochastic regions grow as the perturbation amplitude increases. The stochasticity in the neighbourhood of the hyperbolic fixed points arise due to transverse intersections of stable and unstable manifolds, as already discussed for the driven pendulum example of §1.1.10 (*cf.* Arnold 1963; Walker and Ford 1969; Moser 1973; Lichtenberg and Liebermann 1983, Chapter 3; Guckenheimer and Holmes 1983; Moser 1986; Nicolis 1986b).

Chapter 2

Notation and definitions

It has been found necessary to invoke many technical concepts and to introduce much terminology into this thesis. While much of this is defined throughout the thesis, the argument would be unduly disrupted if every point was to be made precise in the text of each Chapter. Consequently, many of the general mathematical terms and concepts required to provide the proper setting for the other Chapters are defined in this Chapter, so that they can be referred to elsewhere in the thesis. Some general notations are:

$\{\ \}$	a set of
$\cup$	union of sets
$\cap$	intersection of sets
$\subset$	a subset of
$\in$	element of
$\notin$	not an element of
$\emptyset$	empty set
$\|\ \|$	magnitude of
$\dot{x}$	derivative of x wrt time
x'	derivative of x wrt to some variable other than time
$f^{(r)}(x)$	r^{th} derivative of f wrt x
o	composite function
$f^n(x)$	the n-fold composition of f with itself
$\forall$	for all
$\exists$	there exists
iff	if and only if

Useful references for the following sections are incorporated into each section heading.

2.1 Sets (Hausdorff 1957; Chinn and Steenrod 1966; Chillingworth 1976)

A *set* is a collection of mathematical objects. If x is one of the objects belonging to a set X then x is called an *element*, *member* or *point* of X. A set consisting of a single element is called a *singleton*. A set consisting of no elements is called an *empty set*. Some sets of numbers are referred to by the special notations:

$\mathbf{N}$ = the set of all natural numbers = $\{1, 2, \cdots\}$.

$\mathbf{Z}$ = the set of all integers = $\{\cdots, -2, -1, 0, 1, 2, \cdots\}$.

$\mathbf{R}$ = the set of all real numbers.

$\mathbf{R}^+$ = the set of all positive real numbers including 0..

$\mathbf{I}$ = the set of all irrational numbers.

$\mathbf{Q}$ = the set of all rational numbers.

$\mathbf{C}$ = the set of all complex numbers.

$\mathbf{S}$ = the interval of $\mathbf{R}$ from (and including) 0 to (but excluding) $2\pi = [0, 2\pi)$.

Two notations are invoked for specifying particular sets. The first lists the elements of the set between braces (e.g. $\{1, 2, \cdots\}$ is the set of natural numbers). The second notation specifies a set by $\{x : P\}$ and is that set consisting of those elements x having the property P (e.g. $\{x : x \in \mathbf{R}, x^2 = 1\} = \{-1, 1\}$). A *sequence* is a set whose elements are ordered (e.g. the set $\{x_i : i = 1, 2, \cdots\}$, where x_i is the i^{th} element in the sequence). An arbitrary sequence of elements is denoted S_x, where the subscript x identifies the general sequence of elements under consideration. A particular sequence is denoted $S_x(x)$, where the argument in parenthesis identifies the particular sequence under consideration.

An *n-tuple* is denoted $(x_1, x_2, \cdots, x_n)$ and is formed from n objects, where x_1 is called the first *coordinate*, x_2 the second coordinate, etc. The *Cartesian product* of n sets $X_1, X_2, \cdots, X_n$ is defined (and denoted) by

$$X_1 \times X_2 \times \cdots X_n = \{(x_1, x_2, \cdots, x_n) : x_1 \in X_1, x_2 \in X_2, \cdots, x_n \in X_n\} \tag{2.1}$$

and is the set of all n-tuples. The Cartesian product of n identical sets X is written as X^n. The set

$$\mathbf{R}^n = \{(x_1, x_2, \cdots, x_n) : x_1, x_2, \cdots, x_n \in \mathbf{R}\} \tag{2.2}$$

of all n-tuples of real numbers is called n-dimensional *Cartesian space*. An element of $\mathbf{R}$ is called a *scalar*, while an element of $\mathbf{R}^n$ for $n \geq 2$ is called a *vector*. When it is necessary to distinguish between a scalar and a vector, a vector is denoted by a bold face character (e.g. $\mathbf{x} = (x_1, x_2, \cdots, x_n)$) and a scalar is denoted by a normal face character (note that the elements of general sets are denoted by normal face characters). In particular, the n-tuple $\mathbf{R}^1 = \mathbf{R}$ is called the *real line*, and $\mathbf{R}^2 = \mathbf{R} \times \mathbf{R}$ is called the *Cartesian plane*. The algebraic operation called *addition* is defined as

$$\mathbf{x} + \mathbf{y} = (x_1 + y_1, x_2 + y_2, \cdots, x_n + y_n) \tag{2.3}$$

where $\mathbf{x}, \mathbf{y} \in \mathbf{R}^n$ are vectors. The algebraic operation called *scalar multiplication* is defined as

$$\alpha\mathbf{x} = (\alpha x_1, \alpha x_2, \cdots, \alpha x_n) \tag{2.4}$$

where $\mathbf{x} \in \mathbf{R}^n$ and $\alpha \in \mathbf{R}$. A *vector space* is defined as a Cartesian space on which the rules

$$(\mathbf{x} + \mathbf{y}) + \mathbf{z} \quad = \quad \mathbf{x} + (\mathbf{y} + \mathbf{z})$$

$$\begin{aligned} \mathbf{x}+\mathbf{y} &= \mathbf{y}+\mathbf{x} \\ \alpha(\mathbf{x}+\mathbf{y}) &= \alpha\mathbf{x}+\alpha\mathbf{y} \\ \alpha\beta\mathbf{x} &= \alpha(\beta\mathbf{x}) \\ (\alpha+\beta)\mathbf{x} &= \alpha\mathbf{x}+\beta\mathbf{x} \\ \mathbf{0}+\mathbf{x} &= \mathbf{x} \\ \mathbf{x}+(-\mathbf{x}) &= \mathbf{0} \\ 1\mathbf{x} &= \mathbf{x} \end{aligned} \tag{2.5}$$

hold, where $\mathbf{0}$ is the vector having coordinates of all 0's, $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbf{R}^n$ and $\alpha, \beta \in \mathbf{R}$.

A *metric* is a function (refer to §2.3) $d(x, y)$ of two variables which obeys the rules

$$\begin{aligned} d(x,y) &\geq 0 \\ d(x,y) &= 0 \text{ iff } x = y \\ d(x,y) &= d(y,x) \\ d(x,y)+d(y,z) &= d(x,z) \end{aligned} \tag{2.6}$$

where x, y, z are elements from the Cartesian product of n sets X. The set X^n together with the metric is called *metric space*, and is denoted (X^n, d). The metric d is a function from $X^n \times X^n$ to $\mathbf{R}^+$ (i.e. $d : X^n \times X^n \to \mathbf{R}^+$). Cartesian space on which a metric is defined is called *Euclidean space*. For any point $y \in X^n$, the set of all points $\{x : x \in X^n, d(y,x) < \delta\}$ for any arbitrary positive real δ, is termed a *neighbourhood* of y and is denoted N_y. If (X, d_x), (Y, d_y) are two metric spaces, and if $x \in X$, then a function $f : X \to Y$ is said to be *continuous* at x if, for each neighbourhood of $f(x)$ in Y, there is some neighbourhood of x in X whose image is in the neighbourhood of $f(x)$ in Y. If this holds for all $x \in X$ then the function is continuous at each and every point of its domain and is said to be continuous.

$A \subset X$ is called an *open* subset of X if, for each $x \in A$, there exists a $N_x \subset X$ such that $N_x \subset A$. A real number a is said to be the *limit* or *limit point* of the sequence of real numbers $\{a_1, a_2, \cdots\}$ if, given an arbitrary positive real number $\delta > 0$, there is a positive integer n such that $\mid a - a_n \mid < \delta$. If a is a limit point of the sequence $\{a_i : i = 1, 2, \cdots\}$ then for $n \to \infty$ the sequence $\{a_i : i = 1, 2, \cdots\}$ is said to *converge* to a. A set X is closed if and only if, each sequence $\{a_i : i = 1, 2, \cdots\}$ of points from X that converges to a point a, is contained in X. The *closure* of $A \subset X$, denoted $\bar{A}$, is the set consisting of A together with all its limit points (i.e. the set A is closed if and only if $A = \bar{A}$, otherwise $A \subset \bar{A}$). A closed interval of $\mathbf{R}$, $\{x : a \leq x \leq b; a, b, x \in \mathbf{R}\}$ is denoted within square brackets $[a, b]$, an open interval of $\mathbf{R}$ $\{x : a < x < b; a, b, x \in \mathbf{R}\}$ within parentheses (a, b). Two sets A and B are said to be *disjoint* if $A \cap B = \emptyset$. The *complement* of $A \subset X$, is the set of all points in X but not in A. A set X is said to be *countable* if and only if it is possible to associate each element of X with an element of the natural numbers. Otherwise X is said to be *uncountable*.

A collection C of subsets of $\mathbf{R}^n$ is called a *covering* of $X \in \mathbf{R}^n$ if the union of the sets of C contains X (i.e. each point of X lies in at least one of the sets of C). A covering C of X is called *finite* if the number of sets in C is finite. A covering C of X is said to contain a covering D of X if each set of D is also a set of C. A covering of X is called an open covering of X if each set of the covering is an open

subset of X. The space X is called *compact* if each open covering of X contains a finite collection of open subsets of X, whose union is X, and it is possible to select a finite subcollection whose union is also X.

$A \subset X$ is called a *dense* subset, if every point in X is arbitrarily close to points in A. That is, for some metric $d(x, y)$ defined on X, and $y \in A$, there exists a point $x \in X$ such that $d(x, y) < \delta$ for any arbitrary $\delta \in \mathbf{R}^+$.

The *supremum* (sup) or least upper bound (l.u.b.) of $A \subset S$, where the elements of A and S can be ordered, is the smallest element of S, x_ℓ, which is greater than or equal to every element of A. The *infirmum* (inf) or greatest lower bound (g.l.b.) of $A \subset S$, where the elements of A and S can be ordered, is the greatest element of S, x_g, which is less than or equal to every element of A. For example,

$$\begin{aligned} x_g &= \inf(x_1, x_2, \cdots, x_n) = \inf(x_1) \\ x_\ell &= \sup(x_1, x_2, \cdots, x_n) = \sup(x_n) \end{aligned} \tag{2.7}$$

where $x_1, x_2, \cdots, x_n$ is an ordered sequence (i.e. $x_1 < x_2 < \cdots, x_n$) of real numbers.

A *measure* is a mapping μ from a set X to $[0, 1]$ which satisfies the rules,

$$\begin{aligned} \mu(X) &= 1, \\ \text{if } y &= \bigcup_{i=1}^{n} A_i, \text{ then } \mu(y) = \sum_{i=1}^{n} \mu(A_i), \end{aligned} \tag{2.8}$$

where A_i are disjoint sets and $A_i \subset X$. The *Lebesgue measure* $\mu(A)$ is a measure defined on a subset of Euclidean space X. The *length* of an interval is the difference between the supremum of the interval and the infirmum of the interval. In one-dimensional Euclidean space the Lebesgue measure of $A \subset X$ is equal to the length of the interval A divided by the length of the interval X.

Some spaces can be divided in a natural way into two or more parts. For example the complement of a circle in the plane consists of two parts; the part inside and the part outside the circle. Some sets cannot be divided in any natural way. To make this precise, a *separation* of a space X is a pair of non-empty open subsets A, B of X such that $A \cup B = X$, $A \cap B = \emptyset$. A space which has no separation is said to be *connected.* A set is totally *disconnected* if the only connected subsets are single points.

A *Cantor set* is a set which is totally disconnected, compact, uncountable, and has zero Lebesgue measure. As an example, a Cantor set can be constructed by removing sub-intervals from the unit interval $[0, 1]$. Suppose that one removes the middle one-third, then the middle one-third of the remaining two intervals, then the middle one-third of the remaining four intervals, etc. With each step the total length of the interval $[0, 1]$ becomes two-thirds the previous length. The intriguing aspect of the Cantor set is that although the set has no length (in the Lebesgue sense), it contains an uncountable infinity of points (i.e. the Cantor set contains no less points than the original unit interval $[0, 1]$) (*cf.* Dauben 1983).

An *invariant set* Λ for a flow $\phi_t(\cdot)$ on $\mathbf{R}^n$ is a subset $\Lambda \subset \mathbf{R}^n$ such that $\phi_t(x) \in \Lambda$ for $x \in \Lambda$ for all $t \in \mathbf{R}$. The stable and unstable invariant manifolds of a fixed point or periodic orbit are examples of invariant sets. The *nonwandering set*, denoted Ω, is a generalisation of fixed points and periodic orbits. A point p is termed nonwandering for the flow $\phi_t(\cdot)$, if for any neighbourhood of p N_p, there exists an arbitrarily large

t such that $\phi_t(N_p) \cap N_p \neq \emptyset$. The nonwandering set Ω is the set of all such points p. Thus, a nonwandering point lies on or near trajectories (refer to §1.1) which come back within a specified distance of themselves. Fixed points and periodic orbits are nonwandering. Wandering points correspond to transient behaviour, while asymptotic behaviour corresponds to trajectories of nonwandering points.

A point p is said to be an ω-limit point of x if $\phi_t(x) \to p$ as $t \to \infty$. A point q is a α-limit point if $\phi_t(x) \to q$ as $t \to -\infty$. The α- (respectively ω-) limit sets $\alpha(x)$, $\omega(x)$ are the sets of α and ω points of x.

2.2 Linear Algebra (Hirsch and Smale 1974)

This section introduces concepts relating to Cartesian space. A nonempty subset $E \subset \mathbf{R}^n$ is called a subspace of Cartesian space if E is closed under the operations of addition and scalar multiplication in $\mathbf{R}^n$, that is, for all $\mathbf{x} \in E$, $\mathbf{y} \in E$ and $\alpha \in \mathbf{R}$

$$\mathbf{x} + \mathbf{y} \in E, \quad \alpha\mathbf{x} \in E \tag{2.9}$$

where $\mathbf{x}$ and $\mathbf{y}$ are in general vectors. A set of vectors $\{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ in E is said to *span* E if every vector in E is a linear combination of the vectors in the set $\{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$, that is, for every $\mathbf{y} \in E$ there are scalars $a_1, \cdots, a_n$ such that

$$\mathbf{y} = a_1\mathbf{x}_1 + \cdots a_n\mathbf{x}_n \tag{2.10}$$

The set of vectors $\{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ is said to be *independent* if

$$a_1\mathbf{x}_1 + \cdots a_n\mathbf{x}_n = 0 \tag{2.11}$$

only when $a_1 = \cdots = a_n = 0$. A *basis* of E is a set of vectors in E that are independent and spans E. Let $E_1, \cdots, E_r$ be subspaces of E. E is the *direct sum* of them if every vector $\mathbf{x} \in E$ can be expressed uniquely as

$$\mathbf{x} = \mathbf{x}_1 + \cdots + \mathbf{x}_r, \quad \mathbf{x}_i \in E_i, \quad i = 1, \cdots, r \tag{2.12}$$

The direct sum is denoted

$$E = E_1 \oplus \cdots \oplus E_r \tag{2.13}$$

2.3 Calculus (Chillingworth 1976; Guckenheimer and Holmes 1983)

A *map* or *function* is denoted by $f : X \to Y$ and consists of the sets X and Y together with a rule f which assigns to each $x \in X$ a unique element $f(x) = y$ where $y \in Y$. The element y is called the value of f at x, or the *image* of x, and the element x is called the *preimage* of y. The set X is called the *domain* and the set Y the *codomain*. If Y is identical to the set $\{f(x) : x \in X\}$ then Y is called the *range*. The range is a subset of the codomain. If the domain and range are identical then the function is called an *operator*.

A function is said to be *one-to-one* if $f(x) \neq f(y)$ whenever $x \neq y$. The only type of function that can be one-to-one necessarily increases or decreases monotonically. A function $f : X \to Y$ is said to be *onto* if for any $y \in Y$ there is at least one $x \in X$

such that $f(x) = y$. A function is onto if every value in the range of f has at least one corresponding value in the domain. Functions which are one-to-one are also called *injective*, while functions which are onto are also called *surjective*.

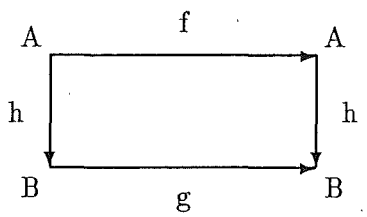
The function $f : \mathbf{R} \to \mathbf{R}$ is differentiable at x if

$$\lim_{h \to 0} \frac{1}{h}(f(x+h) - f(x)) \tag{2.14}$$

exists. This limit is called the *derivative* of f at x, and is denoted $f'(x)$. The r^{th} derivative of f at x is denoted $f^{(r)}(x)$. A function f is of class C^r over an interval I, if $f^{(r)}(x)$ exists and is continuous for all $x \in I$. A function is said to be *smooth* if it is of class C^1. The composition of two functions $f(g(x))$ is denoted by $f \circ g(x)$ and the n-fold composition of f with itself is denoted by $f^n(x)$. Note that f^n does not mean $f(x)$ raised to the n^{th} power, nor does it mean the n^{th} derivative of $f(x)$ which is denoted by $f^{(n)}(x)$.

A function $f : I \to J$ is called a *homomorphism* if f is one to one, onto and continuous, and $f^{-1}(x)$ is also continuous. A function $f : I \to J$ is a C^r diffeomorphism if f and f^{-1} are of class C^r.

Consider two functions $f : A \to A$ and $g : B \to B$. They are said to be *topologically conjugate* or *topologically equivalent* if there exists a homomorphism $h : A \to B$ such that $h \circ f = g \circ h$. The homomorphism h is called a topological conjugacy. In diagrammatic form $h \circ f = g \circ h$ can be represented as



Let $f : I \to I$ and $g : J \to J$ be two functions. The C^0-distance between f and g, denoted $d_0(f,g)$, is given by

$$d_0(f,g) = \sup_{x \in \mathbf{R}} \mid f(x) - g(x) \mid \tag{2.15}$$

The C^r-distance, denoted $d_r(f,g)$, is given by

$$d_r(f,g) = \sup_{x \in \mathbf{R}} \left(\mid f(x) - g(x) \mid, \mid f'(x) - g'(x) \mid, \cdots, \mid f^{(r)}(x) - g^{(r)}(x) \mid \right) \tag{2.16}$$

Two functions are said to be C^r-close or C^r-near over an interval if they as well as their first r derivatives differ by only a small amount. A function $f : J \to J$ is said to be *C^r-structurally* stable on an interval J if there exists an arbitrary $\varepsilon > 0$ such that whenever $d_r(f,g) < \varepsilon$ for $g : J \to J$, f is topologically equivalent to g. Note that, conservative systems (refer to §1.2) are not structurally stable, since the addition of any dissipative term (e.g. a slight friction added to a pendulum) alters qualitatively the state space by introducing attractors.

A *manifold* is a connected surface (or line or hypersurface) having the property that there exists a homomorphism that maps a neighbourhood of every point on the manifold to Euclidean space. Manifolds are usually thought of as smooth surfaces (or lines or hypersurfaces).

A function is said to be *linear* if the following properties hold:

$$\begin{aligned} f(x+y) &= f(x)+f(y) \\ f(\alpha x) &= \alpha f(x) \end{aligned} \tag{2.17}$$

A linear function is of the form $f(x) = ax$, where a is an arbitrary constant. A function is said to be *affine* if $f(x) = ax + b$, and a function is *piecewise-linear* if $f(x)$ is affine over a collection of intervals.

The delta dirac function $\delta(x)$ is defined as

$$\delta(x) = 0 \text{ if } x \neq 0 \tag{2.18}$$

$$\int_{-\infty}^{\infty} \delta(x)dx = 1 \tag{2.19}$$

2.4 Probability (Feller 1966; Ambrozy 1982; Gardiner 1983)

Probability theory is rooted in the real-life situation where a person performs an experiment, the outcome of which may not be certain. Such an experiment is called a *random experiment.* A single performance of a random experiment is called a *trial* for which there is an *outcome.* The totality of possible outcomes of a random experiment is called the *sample space*, denoted S. An *event* is a subset of the sample space. Let A be an event (i.e. $A \subset S$), and let the outcome of a specific trial be denoted s (i.e. $s \in S$). If $s \in A$, then the event A has occurred (e.g. in the experiment "draw a card from a deck of 52 cards", one might only be interested in whether a spade is drawn).

A *random variable* X is a function that assigns a numerical value to each possible outcome of an experiment. The term random variable is a misnomer, since a random variable is a function whose domain is the sample space S, and whose range is the set of real numbers $\mathbf{R}$ (i.e. $X(s) = x$, where $s \in S$, $x \in \mathbf{R}$, and is often abbreviated to $X = x$).

To each event defined on a sample space S, it is possible to assign a non-negative number called *probability.* The probability of event A is denoted $P(A)$. Probability satisfy the three axioms:

$$\begin{aligned} P(A) &\geq 0 \\ P(S) &= 1 \\ P\left(\bigcup_{i=1}^{N} A_i\right) &= \sum_{i=1}^{N} P(A_i) \quad \text{if} \quad A_m \cap A_n = \emptyset \quad \text{when} \quad m \neq n \end{aligned} \tag{2.20}$$

The first two axioms ensure that probability lie somewhere in the interval $[0, 1]$. The third axiom states that if the probability of an event is equal to the union of any number of mutually exclusive events, the probability of the event is equal to the sum of the individual event probabilities.

The probability $P(X \leq x)$ is the probability of the event $(X \leq x)$. It is a number that depends on x and is therefore a function of x. This function is denoted $Q_X(x)$ and is called the *probability distribution function* (CDF) of the random variable X. Thus

$$Q_X(x) = P(X \leq x), \quad -\infty < x < \infty \tag{2.21}$$

The subscript X denotes the random variable under consideration. The *probability density function* (pdf), denoted by $q_x(x)$, is defined as the derivative of the CDF

$$q_X(x) = \frac{dF_X(x)}{dx} \tag{2.22}$$

Chapter 3

Significance of Deterministic Chaos

> Whether or not it is clear to you, no doubt the universe is unfolding as it should (Davies 1989, page vi).

The purpose of this Chapter is threefold: 1) to discuss the relationship between randomness and deterministic chaos, 2) to discuss the implications of deterministic chaos for science and technology and 3) to discuss where experimental observation of deterministic chaos has occurred, and where present experimental research effort is being directed.

§3.1 discusses the relationship between deterministic chaotic dynamical systems and stochastic systems. Deterministic chaotic dynamical systems can to varying degrees mimic stochastic systems. It is possible to classify deterministic systems according to how closely they approximate stochastic systems. §3.1 indicates where deterministic chaos fits into this classification.

The first part of §3.2 discusses some of the wider significance and implications of deterministic chaos for science. The use of computers has enabled nonlinear systems to be intensively studied (Campbell *et al.* 1985). Such studies have revealed that nonlinear systems have entirely different characteristics from linear systems. The organisation and cooperation, for example, of molecules in a biological cell depends on nonlinearity (Coveney 1988; Davies 1989). Such highly organised behaviour, particularly in biology and chemistry, has inspired the study of self-organising (dynamical) systems (this has also become known as synergetics) (*cf.* Pacault and Vidal 1978; Haken 1981; Prigogine and Stengers 1984; Nicolis 1986a). The latter part of §3.2 discusses some technological implications of deterministic chaos. In many industries it is still customary to design and test subsystems in isolation. When these subsystems are brought together to form an entire system, various unexpected or unpredictable phenomena can result, due to system nonlinearities.

§3.3 overviews the experimental observation of deterministic chaos, suggests in what ways it may be of scientific and technological importance, and indicates where current experimental and theoretical effort is being directed.

3.1 Deterministic Chaos and Randomness

The discovery of deterministic chaos has clouded the distinction between determinism and stochasticity, and between predictability and unpredictability. Wolfram (1985) has even suggested that the source of all randomness might be due to deterministic chaos. However, deterministic chaos is usually thought of as being somewhere in between regular deterministic behaviour and truly stochastic behaviour (*cf.* Campbell *et al.* 1983; Chernikov *et al.* 1988).

Each state of a deterministic dynamical system defines a unique history and a unique future for the system (Hirsch and Smale 1974, page 22), which is not the case for a stochastic system. Whatever state a stochastic system is prepared in, the system evolves probabilistically (i.e. it wanders randomly over the allowed region in state space). Moreover, if the same stochastic system is reset to the same initial state, the system evolves entirely differently in general (*cf.* van Kampen 1983; Gardiner 1983). If the state of a deterministic system is either unknown or unknowable it may be indistinguishable from a stochastic system. Deterministic systems which have (extremely) high numbers of states are usually modelled by stochastic systems containing much fewer states. For example, statistical mechanics ignores the states of individual atoms (i.e. their positions, velocities, angular momenta, etc.), but considers the average of atomic states over an extremely large number of atoms (Penrose 1979). These averages form the states of the stochastic process used to model the underlying atomic behaviour. However, it is usually assumed that the actual assemblage of atoms behaves entirely deterministically. Because of the extreme inconvenience involved in deterministic modelling of such an assemblage, one is forced to fall back on probabilistic approaches such as statistical mechanics. Incidentally, the direct demonstration of the underlying deterministic behaviour of individual atoms resulting in the statistical mechanical model has yet to be confirmed (*cf.* Campbell *et al.* 1985).

According to May (1976), writing before the current wide appreciation of deterministic chaos, it was thought that a low-dimensional deterministic system could (or would) not behave like a stochastic process. Furthermore, it was thought that the best way to model low-dimensional deterministic systems was with deterministic equations. However, there now exist many examples of low-dimensional (i.e. of three dimensions) systems that behave to various approximations like stochastic systems. It was also generally thought that stochastic systems provided the most convenient (or best) way to model extremely high-dimensional deterministic systems. Deterministic chaos has blurred our previous definite ideas on the type of model (i.e. stochastic or deterministic) which may be most suitable in any particular situation.

Osborne *et al's.* (1986) study of the motion of a buoy in the Pacific Ocean provides an informative illustration of the search for a low-dimensional deterministic system to describe an apparently stochastic system. The purpose of the study was to determine if the buoy's motion was described by a low-dimensional deterministic system rather than a high-dimensional deterministic or stochastic system. The conclusion of the study was that while the analysis ruled out the possibility of a low-dimensional attractor, it was not in general possible to decide whether the data were characterised by stochastic behaviour or by deterministic dynamics with a large number of degrees of freedom.

A deterministic system can only ever approximate a (truly) stochastic system (*cf.* Lebowitz and Penrose 1973; Penrose 1979; Lichtenberg and Lieberman 1983). However, there exists a hierarchy of properties characterising how random the behaviour of a system is (each property implying the preceding one). A system possessing only properties near the bottom of this hierarchy acts little like a stochastic system, while a system possessing properties near the top of this hierarchy acts most like a stochastic system. Three such properties, which are discussed below, are termed: ergodic, mixing and Bernoulli. The concept of ergodicity is invoked in Chapter 5. A Bernoulli system is mixing and ergodic, while a mixing system is ergodic.

A dynamical system is said to possess ergodicity if the trajectory of a state returns arbitrarily close to that state an infinite number of times as $t \rightarrow \infty$. An analogy to ergodicity can be drawn with a walk in a snow covered park. The whole park will eventually become covered with footprints so that one is eventually forced to walk on previous footprints over and over again. For any ergodic system there exists a probability density function (pdf) $q_X(x)$, defined in the limit $t \rightarrow \infty$, for the time a trajectory spends in any subset x of the allowed region of state space, normalized by the time since the system was set to its initial conditions (*cf.* Lebowitz and Penrose 1973; Eckmann and Ruelle 1985). This pdf is termed the *invariant measure* or the probability distribution. The pdf $q_X(x)$ is actually invariant because it does not change under the flow $\phi_t(x)$ of the dynamical system (i.e. $q_X(x) = \phi_t(q_X(x))$, as indicated in §1.1). The time average of any *observable function* $f(x)$ of the state variables is written as

$$< f(x) >= \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f(\phi_t(x)) dt \tag{3.1}$$

The spatial average over the same observable function is written as

$$\bar{f}(x) = \int_S f(x) q_X(x) dx \tag{3.2}$$

where S denotes the allowed region of state space. Since the trajectories of an ergodic system wander over all the allowed region of state space, the time average of any observable function equals the spatial average over the same observable function (Lebowitz and Penrose 1973). Thus, the consequence of ergodicity is that

$$< f(x) >= \bar{f}(x) \tag{3.3}$$

Consider, as an example, what has come to be known as the twist map of a circle (Lichtenberg and Lieberman 1983, page 262). It is defined by

$$\theta_{n+1} = (\theta_n + 2\pi\alpha) \bmod(2\pi) \tag{3.4}$$

where $(a) \bmod(b)$ returns the remainder of a/b, θ_n is the modulo 2π angular displacement around the circle at the n^{th} instance and α is an arbitrary real constant lying in the interval $(0, 1)$. The map is ergodic over the circle if α is an irrational number, because then the trajectory of the map densely covers the circle (i.e. the trajectory does not form a periodic orbit) (Lichtenberg and Lieberman 1983, page 260). If α is a rational number (i.e. $\alpha = p/q$ where p and q are integers) the twist map is not ergodic over the circle, since the trajectory of the map is periodic with period q/p (i.e. the trajectory does not densely cover the circle). All dynamical systems are ergodic over some subset of state space. For example, the twist map for

α rational, although not ergodic over the circle, is ergodic over the subset of points $\{\theta_i = (\theta_0 + 2\pi pi/q) \bmod(2\pi) : i = 1, \cdots, q/p\}$. Thus, it seems that the central question is to define over what subset of state space a dynamical system is ergodic (although in statistical mechanics the subset of state space is taken to be the region which is consistent with the total energy contained in the system).

A dynamical system is said to possess mixing if

$$\lim_{t\to\infty} q_X[\phi_t(A) \cap B] = q_X(A)q_X(B) \tag{3.5}$$

where $q_X(\cdot)$ is the invariant measure, and A and B are subsets of the allowed region of state space (Lebowitz and Penrose 1973, page 26). A mixing system distorts any region of state space so strongly that filaments of the region eventually spread over the whole of the allowed region of state space, just as a drop of milk (the volume occupied by the drop corresponds to a subset of the allowed region of state space) is distributed throughout a glass of water after it has been stirred. For example, the twist map is not mixing, since any arc on the circle (i.e. subset of the circle) is not eventually distributed around the circle.

A Bernoulli dynamical system is equivalent to the spinning of a roulette wheel (Lebowitz and Penrose 1973, page 28). A finite partition of the allowed region of state space is defined as any finite collection of n nonoverlapping regions (i.e. $R_0, \cdots, R_{n-1}$) which together cover the allowed region in state space. Suppose some measuring instrument could determine which of these regions the trajectory is in at any time, with each region labelled with a different positive integer. Each time the measuring instrument is used it generates a positive integer labelling the region which the trajectory is in at that time. Suppose the instrument is used repetitively, thereby generating a sequence of positive integers. In general, these integers are correlated. However, for some deterministic dynamical systems it is possible to choose the regions $R_0, \cdots, R_{n-1}$ in such a way that the measurements made at different times are completely uncorrelated, just like an ideal roulette wheel (i.e. effectively transforming the dynamical system to a symbolic dynamical system; refer to §1.1.7). Thus a deterministic dynamical system can show determinism on a microscopic level (i.e. the trajectory) but stochasticity simultaneously on a macroscopic level (i.e. regions of state space).

It is not possible to decide if a sequence of numbers originates from a deterministic or stochastic source (Chaitin 1975). However, it is possible to define how random a sequence of numbers is by determining the amount of information, as defined by Shannon (1948), it contains. According to Chaitin (1975) the information embodied in a random sequence of numbers cannot be compressed or made more compact. A definition for randomness based on this idea, known as the algorithmic definition of randomness, states that a sequence is random only if it is impossible to devise an algorithm (for generating the sequence) which is shorter than an algorithm which merely reads back the sequence. This idea can be put on a formal basis by defining the minimal computer program size K_C required to generate an arbitrary sequence S_N of N bits (Lichtenberg and Lieberman 1983, page 274). K_C is the minimum number of bits required to instruct a computer C to generate the N bits of the sequence. This definition depends on C, but it can be decomposed into a machine independent and a machine dependent part, i.e.

$$K_C = K_N(S_N) + C_C \tag{3.6}$$

where C_C depends on C but is independent of S_N, and K_N is independent of C. The lower bound for K_N is unity (in which case the sequence is entirely regular and can be described by one bit of information), the upper bound for K_N is N (in which case the sequence cannot be compressed). When $K_N = N$ the sequence is said to be random. The values of K_N for sequences of numbers generated by deterministic systems can range from 1 to N. A Bernoulli deterministic system can generate a sequence of numbers for which $K_N = N$, whereas $K = 1$ for a deterministic system operating at a fixed point. According to Lichtenberg and Lieberman (1983, page 175), it can be proved that almost all sequences of numbers are random in this sense (i.e. the set of nonrandom sequences is of zero measure).

The algorithmic definition of randomness has implications for number theory. A *computable number* is a number whose decimal expansion can be computed to an arbitrarily large precision via an algorithm requiring a finite amount of information to implement. Any irrational number, such as $\sqrt{2}$, π or e, whose digits can be computed via continued fraction expansion is an example of a computable irrational number. According to McCauley (1988, page 51), Turing (1937) proves that computable numbers are countable. This means that almost all irrationals that can be defined (irrationals can be defined as the limit of a sequence of rationals) cannot be computed by any possible algorithm that can ever be invented (i.e. decimal expansions of such numbers contain infinite amounts of information). This implies the decimal expansion of almost all numbers are completely unpredictable or random in the algorithmic sense.

Consider as an example the one-dimensional shift map defined on the interval $[0, 1)$:

$$x_{n+1} = (10x_n) \bmod(1) \tag{3.7}$$

with the initial condition (i.e. the first member of the sequence) lying in the interval $[0, 1)$ (i.e. $0 \leq x_0 < 1$). Note that, each iteration of the shift map (3.7) stretches every subset of the interval $[0, 1)$ (the stretching operation is performed by the multiplication by 10) and folds back into $[0, 1)$ any subset that has escaped $[0, 1)$ (the folding operation is performed by the modulo operation). It is convenient to define the finite collection of nonoverlapping regions $R_0, \cdots R_{n-1}$, to be the ten equal length intervals $[0, 0.1), \cdots, [0.9, 1.0)$. When $x_0 = 0.29734516\cdots$, the shift map (3.7) generates the sequence

$$\{a_i : i = 0, 1\cdots\} = \{2, 9, 7, 3, 4, 5, 1, 6, \cdots\}. \tag{3.8}$$

Whether or not this sequence is random, hinges on whether the decimal expansion of the initial condition x_0 is random. Since decimal expansions of almost all numbers (including the initial condition x_0 of (3.7)) are random, so are the dynamics of (3.7), which corresponds to a Bernoulli system. Ford (1983) argues that (3.7) is as random as a coin toss. However, this does not imply that deterministic chaos depends solely upon the random character of the initial condition x_0. While the trajectory of a deterministic system is at liberty to depend sensitively upon the random character of the initial condition, it is not required to do so. Also it has not been demonstrated that this is the cause of the deterministic chaos that arises in the majority of chaotic systems, although according to Lichtenberg and Lieberman (1983, page 275) it has been conjectured that the motion near transverse homoclinic or heteroclinic points is random in this sense.

Recall, from the second paragraph of this section, that the entire history and future of a deterministic dynamical system is characterised (i.e. specified uniquely) by any point (i.e. initial condition) on its trajectory. Each and every point on the trajectory contains all the information necessary to fully characterise the trajectory. Therefore, all the points in state space passed through by a trajectory of a deterministic system exhibiting random behaviour (in the algorithmic sense) must be characterised by numbers that are uncomputable (provided the system equations themselves do not contain the infinite information required to generate such behaviour). The equations characterising such behaviour require all the information contained in any point (which are characterised by uncomputable numbers) on the trajectory to fully specify the trajectory. This implies that the trajectory depends sensitively on its initial conditions. Two uncomputable numbers (i.e. points in state space), no matter how close together, contain almost no information that is common, and therefore characterise completely different trajectories. Thus for example, if such a trajectory characterises the motion of a particle, then its motion is random and is completely unpredictable or uncomputable. A chaotic dynamical system exposes to the world the information contained in a number (i.e. a number characterising a point in state space), by reading out serially the information contained in the decimal expansion of that number (starting from the beginning of the expansion and progressing down through the expansion), through stretching and folding as demonstrated by (3.7). On the other hand, a non-chaotic dynamical system does not in general require all the information contained in any point of state space to fully specify its trajectory. Thus, the trajectory of a non-chaotic system can pass through points in state space characterised by both computable and uncomputable numbers.

If a dynamical equation is to be capable of describing observation, one must take into account that absolute measurement precision (which requires infinite information) is unachievable. According to Coveney (1988) this assertion has far-reaching implications. The concept of a point in state space must be replaced by the concept of a region (corresponding to the measurement uncertainty) in state space. This implies that, for a certain class of dynamical system (i.e. chaotic dynamical systems), the whole concept of trajectories becomes operationally meaningless, and even the idea of classical determinism becomes questionable.

Wolfram (1975) identifies two categories of deterministic chaotic dynamical system: 1) the seemingly random behaviour of the system is introduced through the initial condition, for example, the shift map described by (3.7) – Wolfram (1975) calls such system behaviour homoplectic; 2) the seemingly random behaviour of the system is intrinsic and is not totally dependent on the initial conditions, for example, the twist map described by (3.4) where α is an uncomputable number (in this case the system equation contains the infinite information required for general random behaviour) – Wolfram (1975) calls such system behaviour autoplectic. Wolfram's categorisation has not been widely adopted in the literature, although it does seem useful, since it identifies two fundamentally different characteristics of chaotic behaviour. This lack of acceptance of Wolfram's nomenclature is perhaps due to most systems not yet being understood well enough to convince people that such a categorisation is really appropriate.

3.2 Scientific and Technological Significance of Deterministic Chaos

Science has traditionally attempted to explain all physical reality in terms of a few underlying fundamental principles (i.e. the activity of fundamental particles and fields) (Davies 1989). According to this philosophy, biology will ultimately explain sociology and psychology. Chemistry will explain biology and fundamental physics will explain chemistry. Thus, ultimately, physical reality depends solely on the activity of elementary particles and fields. The point of contact between the sciences is at the bottom or fundamental level only, and the only true laws are the laws of particle physics. This is known as the *reductionist approach* to science: all phenomena can be explained in terms of, or reduced to, a few universal and fundamental principles. However, according to Davies (1987) and Crutchfield *et al.* (1986), there seems to be an increasing awareness that the reductionist approach is perhaps naive. Prigogine (1980,1984) points out that questions such as the following seem to be difficult to answer satisfactorily:

- How can the complexity of galaxies, life, etc., which appear to be evolving from disorder to order, be reconciled with the second law of thermodynamics?
- How do billions of atoms cooperate together in chemical oscillators, lasers, etc., in ways which would seem impossible if viewed merely on an atomic scale?
- How can the apparently random nature of such things as some moons of planets (e.g. Saturn's moon Hyperion), dripping taps, etc., be reconciled with the underlying deterministic laws of such phenomena?

Popper and Eccles (1977, page 61) have said in relation to the complexity of nature, that "the greatest riddle of cosmology may well be ...that the universe is, in a sense, creative". Concerning the lack of knowledge about the behaviour of many simple systems of equations, such as, the simplified system of equations describing atmospheric turbulence first studied by Lorenz (1963), Hirsch (1984, page 40) comments, "considering all the attention paid to Lorenz's system, this situation is something of a scandal".

The reductionist approach has been superbly successful for dealing with those systems which are linear or nearly linear. In reality all physical systems are nonlinear, but often they behave in an approximately linear way when close to equilibrium (e.g. a pendulum oscillating with a small swing, a liquid in thermal equilibrium, electromagnetism in isolation from interaction with matter, etc.). When a system is not in equilibrium it exhibits in general nonlinearity, so that superposition no longer holds (Coveney 1988). Component parts of a nonlinear system operate differently when isolated from the system, compared to when the same components are part of the complete system. According to Davies (1987,1988) "by focusing attention on linear systems it has become generally accepted that simple equations could (or would) only lead to simple behaviour ...nonlinearity was harder to study but no significantly different behaviour or approach to studying such systems was thought to be required". Examples of systems that cannot be easily interpreted by the reductionist approach arise particularly in the life sciences. For example, the operation of biological cells, the development of an embryo, etc., are systems that contain processes which do not

operate in equilibrium. It appears that such systems cannot be analysed by breaking them down into component parts. The system must be treated as a whole. When a system is driven (by some process external to the system) away from equilibrium (i.e. when nonlinearity becomes significant), the system is liable to leap abruptly and spontaneously into new, more complicated or highly organised states (Prigogine and Stengers 1984).

According to the second law of thermodynamics, the universe is irreversibly evolving towards a state of maximum entropy or maximum disorder. However, the universe also appears to be progressing from featurelessness to states of greater organisation and complication. Consider as an example, a chemical clock (Winfree 1974). According to Prigogine (1980) if any single event has motivated a change in thinking, the discovery of chemical clocks has. The classical theory of chemical kinetics is based on the assumption that the rate of a chemical reaction is proportional to the concentration of the products taking part in it (Prigogine and Stengers 1984, page 133). It is through collisions between molecules that a reaction takes place, and it is assumed that the number of collisions is proportional to the product of the concentrations of the reacting molecules. Chemical kinetics deals with changes in the concentration of the different products involved in a reaction, as is illustrated by the simple reaction equation

$$A + X \rightarrow B + Y \tag{3.9}$$

where A, B, X and Y represent single molecules of different chemicals. This reaction equation states that whenever a molecule of A encounters a molecule of X, there is a certain probability that a reaction takes place and a molecule of B and a molecule of Y is produced. The rate of change $\dot{X}$ of concentration of X, is proportional (by a factor k) to the product of the concentrations of A and X in the solution, likewise the rate of change $\dot{Y}$ of concentration of Y, is proportional (by a factor k') to the product of the concentrations of B and Y in the solution, i.e.

$$\dot{X} = -kAX \quad \text{and} \quad \dot{Y} = -k'BY \tag{3.10}$$

The proportionality factors k and k' are linked to quantities such as temperature and pressure. Whenever a molecule of X disappears a molecule of A disappears too, and a molecule of Y and one of B is formed. The rates of change of the concentrates are related: $\dot{X} = \dot{A} = -\dot{Y} = -\dot{B}$. If left to itself, the system tends toward a chemical equilibrium (or, in other words, an attractor). At equilibrium $\dot{X} = 0$, which implies $AX/BY = k'/k = K$. This result is known as the law of mass action, or Guldberg and Waage's law (Prigogine and Stengers 1984, page 133), and K is the equilibrium constant.

The rates of chemical reactions are also affected by catalysts. An important type of catalysis, particularly in biology, is one in which the presence of the product is required for its own synthesis (Winfree 1974). An important feature of such systems is that the kinetic equations describing the changes occurring in them are nonlinear differential equations. Another important class of catalytic reactions, occurring in biology, is that of crosscatalysis, for example

$$\begin{aligned} A &\rightarrow X, \\ 2X + Y &\rightarrow 3X, \\ B + X &\rightarrow Y + D, \\ X &\rightarrow E \end{aligned} \tag{3.11}$$

where the concentration of the chemical products A, B, D and E are maintained constant (by controlled inflow and/or outflow of the chemicals into the reaction), whereas the two products X and Y may have concentrations that change with time. This reaction was developed at Brussels during the past decade, and the model is called the Brusselator (Prigogine 1980, page 109). If the concentration of reactant B is used as a parameter, the reaction proceeds towards an equilibrium point, provided the concentration of B is below some critical value. Above this value the equilibrium point becomes unstable and a limit cycle forms. Instead of remaining stationary, the concentrations of X and Y now oscillate at a well defined frequency. This is known as a chemical clock. Prigogine and Stengers (1984, page 230) comment that "such a degree of order stemming from the activity of billions of molecules seems incredible, and indeed, if chemical clocks had not been observed, no one would believe that such a process was possible ...such phenomena contradict perhaps the spirit, but not the letter of the second law of thermodynamics ". If the concentration of reactant B is increased further, eventually the limit cycle disappears and is replaced by deterministic chaos.

According to Davies (1989, page 76), there is an important distinction to be made between order and organisation: order refers to the quantity of information required to describe the state of a system (for example, to describe a system in thermal equilibrium requires no information because the system is completely disordered), while organisation refers to the quality of the information used to describe the state of a system. For example, to specify a satellite orbit, one could simply specify the position and velocity of the satellite at some instant or one could specify a list of successive positions of the satellite separated by specified time intervals. Both specifications contain the same amount of information, but the later specification would usually be more useful (i.e. of greater quality). According to these definitions, deterministic chaos is a manifestation of order (since it arises in systems not in equilibrium), but deterministic chaos is disorganised. A chemical clock, on the other hand, is in an organised state of order. Note that, a system in thermal equilibrium is disordered, but can be either organised or disorganised (e.g. water can be in a state of thermal equilibrium when it is in the form of either steam or ice, but steam is disorganised, while ice is organised).

Complicated biological or man-made technological systems are usually modelled by networks or arrays of interconnected nodes or subsystems. These subsystems could represent neurons in the brain, individual logic components in a computer, or switching nodes in a communication network. It is customary to design, develop and test subsystems of a technological system in isolation from the complete system. However, if an analogy can be made between technological systems and systems that are known to self-organise (e.g. the molecules that form robust coherent or chaotic patterns of activity, such as in the Brusselator) then, when subsystems of technological systems are brought together to form entire systems, various unexpected or heretofore unpredictable phenomena may result (due to system nonlinearities). This has serious implications for the reliability and stability of technological systems.

The activity of a system is holistic, it is the whole pattern of activity that is of interest, not the state of any specific subsystem (i.e. it is not possible to reduce the system, or a pattern of activity of the system, to the activity of individual subsystems). To analyse and take advantage of self-organisation in technological systems requires the development of new design and analysis techniques. For example,

each node of a communication network implements algorithms for controlling traffic flow: these algorithms incorporate nonlinearity and many feedback paths. It may eventually prove feasible to design a communication network to operate in a specific self-organising coherent or chaotic way to enhance network performance and improve network fault tolerance.

According to Davies (1987) it may not prove an overly formidable task to discover organisational principles having application to a wide range of systems. However, little is known about organisational principles at present, while the approach needed to develop such principles is not clear. Nevertheless, the spinoffs from such a discovery may turn out to be revolutionary.

3.3 Important Experimental Observations of Chaos

This section discusses a number of experimental observations in which deterministic chaos either has been shown to occur or is thought to have occurred. In almost every case the chaotic behaviour is poorly understood. Most effort has been concentrated on attempting to fully understand the very simplest of systems. Although many of the experimental studies are highly idealised, compared to what transpires in the real-world, the hope is that fuller understanding of these simple systems will give general insight into more complicated systems (Hao 1984, page 66).

The study of chaos is finding applications in a wide variety of physical situations. The list is already large (Berry *et al.* 1987; Hao 1984) including problems in the onset of turbulence in fluids (Lorenz 1963; Ruelle and Takens 1971; Normand *et al.* 1977), chemically reacting systems (Winfree 1974; Prigogine 1980; Tomita 1982), solid state physics (Bak 1982), lasers (Arecchi and Harrison 1987; Harrison 1988), particle accelerators (Percival and Richards 1982), biological population dynamics (May 1976; Guckenheimer *et al.* 1977), electronic circuits (Chua 1987), astromony (Wisdom 1987), numerical methods (Newell 1977; Wong 1984; Kloeden and Mees 1985), physiological systems (Glass *et al.* 1987) and economics (Barnett and Chen 1986; Mason *et al.* 1986). The subsections §3.3.1 through §3.3.5 below briefly describe selected highlights from five of these fields. The purpose of these five subsections is merely to give an indication of the direction of current research activity in leading laboratories. It is not meant to be an exhaustive review.

3.3.1 Electronic Circuits

There has been a wide variety of electronic circuits exhibiting chaotic behaviour reported in the literature (*cf.* Saito 1985; Matsumoto *et al.* 1986b; Kennedy and Chua 1986; Pei *et al.* 1986; McGonigal and Elmasry 1987; Chua 1987; Kuo 1988). Further, the number of reports of chaotic behaviour of relevance to the technological sciences, seems to be accelerating (*cf.* Esande 1985, page 168). For example, note the number of papers on chaos published in the IEEE Transactions on Circuits and Systems in 1988 as compared to previous years.

The most widely reported and studied electronic circuit that can exhibit chaotic behaviour is a particular type of oscillator (Chua *et al.* 1986), now referred to as Chua's circuit. This consists of one nonlinear negative resistance, and three energy storage elements. It is the simplest autonomous circuit which can become chaotic.

It is unique in that it is the only known example of a physical system which has been shown to be chaotic using the three different approaches: computer simulation, laboratory experiment and mathematical analysis (refer to §6.6). Summarised below are other reports of electronic circuits exhibiting deterministic chaos.

Digital filters. Chua and Lin (1988) have shown that chaotic behaviour can occur in 2nd order finite impulse response (FIR) digital filters. An adder is a circuit element essential for almost all digital signal processing applications. The numerical overflow from an adder in a digital filter feedback loop can introduce sufficient nonlinearity to cause deterministic chaos. Chua and Lin (1988) show under what conditions such a filter can become chaotic. They conclude that urgent research is required into this phenomenon, because it may have important implications for digital signal processing in general.

Control systems. Ushio and Hsu (1987) have shown that roundoff error in a combination digital-analogue control system can cause instabilities and chaotic behaviour. They suggest that this may have significant implications, since control systems are becoming more complicated and inherently digital in character. Sparrow (1980), Brockett (1982) and Ushio and Hirai (1985) have studied simple nonlinear feedback systems having piecewise linear characteristics. They have demonstrated some sufficient conditions for deterministic chaos to occur in such systems. However, they reach few concrete conclusions but emphasize that much further work is required.

Josephson junction. The Josephson junction is a superconducting tunnel diode junction (Pederson *et al.* 1973; Salam and Sastry 1985). According to Hao (1984, page 71), noise power can grow anomalously with gain level for certain parametric amplifiers incorporating a Josephson junction. The equivalent noise temperature can be as high as 50,000K, even though the actual temperature of the device is maintained at 4K. Such a high noise level cannot be explained by any known natural noise source. Huberman *et al.* (1980) formulates equations for the dynamics of the junction, which when numerically modelled predict deterministic chaos (also confirmed experimentally).

Oscillator synchronization. Synchronizing an oscillator to an external signal can lead to a phenomenon known as the devil's staircase, as well as to deterministic chaos. If the frequency of the external signal is slowly increased, the synchronized oscillator frequency tends to increase in discrete jumps (Bak 1986; Glazier and Libchaber 1988). The totality of these jumps forms the devil's staircase (a staircase that supposedly only the devil could climb!). The study of driven oscillators is of considerable importance since many physical phenomena can be modelled this way (e.g. lasers, physiological processes, synchronisation of electronic oscillators, etc.). The most studied driven oscillators are of the van der Pol and the Duffing varieties (Chandra and Scott 1981). Jefferies (1986) shows how a simple monostable oscillator triggered from an external periodic source, can become chaotic. He concludes that autonomous timing elements should not be included in clocked digital systems.

Switched Capacitor Circuits. Rodriguez-Vazquez *et al.* (1985) have demonstrated that a particular switched capacitor circuit is equivalent to the logistic

map (refer to §1.1.6). They conclude that the dynamics of all discrete time circuits should be carefully studied before they are implemented in VLSI.

3.3.2 Physiological Systems

In the human body there are many feedback mechanisms working to maintain physiological variables within "normal" limits. Many physiological variables display complicated temporal fluctuations (even when the environment is maintained as invariable as possible) which are difficult to characterise (Glass *et al.* 1987). Goldberger *et al.* (1985, 1987, 1988) has suggested that these fluctuations may be the result of chaotic dynamics. Goldberger *et al.* further claim that pathological conditions may result when such fluctuations become regular (i.e. a bifurcation out of chaotic behaviour into regular behaviour).

Much research has centred on cardiovascular dynamics. A healthy heartbeat is not perfectly regular and it fluctuates in a highly erratic fashion, with an excess low-frequency power spectrum (refer to Chapter 5). The result of constructing a state space from such fluctuations suggests the presence of a strange attractor (Goldberger and Rigney 1988). Preliminary data indicate that pathologic heartrate dynamics have less variability and less low frequency energy content than those of a healthy heart. Convincing evidence for chaotic dynamics in normal physiological functions still seems to be lacking, but several theoretical and experimental studies have pointed to possible causal mechanisms for chaotic dynamics in physiological systems (*cf.* Glass *et al.* 1983; Chialvo and Jalife 1987; Babloyantz and Destexhe 1988; Glass *et al.* 1988). Two proposed causes of chaos are: time delays within physiological control feedback loops, and synchronization of self oscillating physiological systems to external periodic stimulations.

It is known that control systems incorporating time delay, and positive and negative feedback, are capable of chaotic dynamics (Sparrow 1980). Glass *et al.* (1988) model physiological control feedback by

$$\dot{x}(t) = f(x(t - T_d)) - g(x(t - T_d))x(t) \tag{3.12}$$

where $x(t)$ is the variable to be controlled, T_d is the duration of the delay, $f(\cdot)$ is a function describing the production rate of $x(t)$, and $g(\cdot)$ is a function describing the decay rate of $x(t)$. Glass *et al.* (1988) and Farmer (1982) have invoked (3.12) for a model of blood production, in which

$$f(\cdot) = \frac{ax(t - T_d)}{1 + x^c(t + T_d)} \quad \text{and} \quad g(\cdot) = b \tag{3.13}$$

Farmer (1982) demonstrates that for particular values of a, b, c and T_d, the dynamics of $x(t)$ are chaotic. Glass *et al.* (1988) concludes that "fluctuation (in physiological variables) may be due, at least partially to chaotic dynamics which arise as a consequence of ...feedback mechanisms in nonlinear physiological control systems."

Cardiac arrhythmias are abnormal cardiac rhythms (some of which are life-threatening, but most of which are not) of the human heart. Glass *et al.* (1983) have developed a model which characterises the normal heart rate regulation, and some cardiac arrhythmias, of an aggregate of spontaneously beating cells (from an embryonic chick heart), which are controlled through external periodic stimulation.

The dynamics of the cells depend on the repetition rate M and the strength of the external stimulations. Under certain conditions the dynamics of the cells become periodic (i.e. repeat after N external stimuli). This is known as $N : M$ phase locking, frequency locking or resonance (Bak 1986; Glazier and Libchaber 1988). Under other conditions, the dynamics of the cells appears to be chaotic. Glass *et al.* (1983) claim that there is strong agreement between their experimental results and their model (a non-invertible one-dimensional circle map). They conclude that their work may give a much deeper understanding of the dynamics of types of arrhythmia which are often encounted clinically.

3.3.3 Hydrodynamic Turbulence

As Swinney (1978) explains, laser doppler velocimetry and modern data acquisition techniques have allowed accurate experiments to be performed on hydrodynamical systems. Many experiments have concentrated on the onset of turbulence in finite containers. Bénard instability (thermoconvective instability of fluid in a container heated from below with a free upper surface), Rayleigh-Bénard instability (Bénard instability with an upper plate), and Couette flow instability (flow between rotating cylinders) are types of finite container instabilities that have been intensively studied (*cf.* Normand *et al.* 1977; Dold and Eckmann 1977; Eckmann 1981). As regards shear flow turbulence (which is perhaps more important), knowledge of chaos has helped little, because infinite degrees of freedom seem to be involved.

Landau (1959, Chapter 3) suggests that an infinite sequence of instabilities should occur at the onset of turbulence, each instability adding a new frequency to the motion. Turbulence would then be a motion consisting of a superposition of so many frequencies that it would be too complicated and unstructured to permit other than a statistical characterisation. Ruelle and Takens (1971) have pointed out that the transition to turbulence is more likely to occur, for most flows, following only three or four instabilities, the resulting flow being chaotic. Few experiments have succeeded in distinguishing between chaotic flow and complicated quasiperiodic flows. However, it appears that flows generally become chaotic after a small number of instabilities, as Ruelle and Takens suggest, but there are substantial differences in detail marking the transition process to turbulence in different hydrodynamical systems.

3.3.4 Optical Turbulence

According to Harrison and Biswas (1986) and Harrison (1988), the recent discovery that optical systems (in particular lasers) can exhibit deterministic chaos is particularly significant, because such systems are ideally suited to experimental investigation. Optical instabilities occur over very short time scales (nanoseconds to microseconds) enabling essentially constant environmental conditions to be maintained. This is of practical importance, because extraneous environmental perturbations (such as temperature fluctuations, noise, etc.) can dramatically alter the form of the temporal evolution of any chaotic process, making quantitative analysis difficult. Also, optical systems are relatively easy to construct, and are characterised by relatively simple mathematics.

Chaotic behaviour has been observed in two optical areas: lasers or active systems (in which the optical signal is derived from stimulated emission generated within an

optical cavity) and passive systems (in which an optical signal is transmitted through an optical cavity containing, for example, a nonlinear medium which is modified by the intensity of the optical signal).

According to Ackerhalt *et al.* (1985, page 246) it is not unusual for lasers to have noisy or even unstable outputs. Such behaviour is usually attributed to certain random factors such as cavity impurities, defects, etc. Recently, however, it has been discovered that a laser output can be chaotic. The Maxwell-Bloch equations (*cf.* Graham 1984; Arecchi and Harrison 1987) describe a unimodal laser operating under special conditions. Haken (1975) has shown that these equations can be transformed into the simplified system of equations describing atmospheric turbulence introduced by Lorenz (1963), which have been extensively studied and known to exhibit deterministic chaos. As the excitation energy is increased the output from such a laser exhibits a series of bifurcations culminating in chaotic behaviour. Since Haken's discovery, many experimentalists have reported chaotic behaviour in a wide range of different lasers (*cf.* Graham 1984; Ackerhalt *et al.* 1985; Arecchi and Harrison 1987).

A passive optical system transmits optical signals only. Optical instability can occur in such systems if the absorption coefficient and/or the refractive index of the transmission medium is modified by the intensity of the optical signal, or feedback loops are involved. Passive systems are being increasingly recognized for their potential implementation as optical logic elements. Instabilities in such systems need to be avoided. Nevertheless, it may be possible to take advantage of the periodic instabilities that precede chaos in the development of, for example, ultra-high frequency optical modulators.

Passive optical systems that have been much studied are optical cavities that incorporate time delay feedback loops. These systems are often bistable, in the sense that any such system has two stable output states for the same input state. Ikeda (1979) was the first to predict theoretically that bifurcations and chaos may appear in such devices. This effect has been observed experimentally in a hybrid bistable device (i.e. part of the optical output is delayed by an electronic circuit and then fed back to the input) (Arecchi and Harrison 1987). Such devices can be described by ordinary differential equations incorporating time delays. According to Hao (1984), hybrid bistable devices usually incorporate feedback time delays of the order of milliseconds, so that the structure of this chaotic behaviour has little to do with the underlying optics. Much effort has been expended on the search for purely optical chaos, with many successes being recently reported, such as the Q-switched laser (Arecchi *et al.* 1982).

3.3.5 Astronomical Chaos

Astronomy has revealed a rich source of chaotic behaviour (Wisdom 1987; Lerche and Low 1982). The rotation of Saturn's satellite Hyperion is currently chaotic, which is primarily a consequence of Hyperion's highly aspherical shape (Binzel *et al.* 1986). According to Wisdom (1987a, 1987b), many of the other irregularly shaped satellites in the solar system have had chaotic rotations in the past. After millions of years, tidal forces pull the satellites into regular motion. Wisdom (1987a, 1987b) also claims that meteorites are most probably transported to Earth from the asteroid belt by way of chaotic orbits.

The Kirkwood gaps in the asteroid belt have been an object of a great deal of speculation (Wisdom 1982). The basic difficulty in understanding the formation of the gaps had been that the orbits of the asteroids are not well understood analytically, and numerical simulations have not alleviated the problem because of the great amount of computer time required. Long numerical integrations have been recently achieved by a new method for following the trajectories of asteroids (Wisdom 1987). These results suggest that asteroids, after spending a hundred thousand years or longer in a low eccentricity orbit, can suddenly take large excursions or burst into an irregular high-eccentricity orbit, returning relativity quickly back to a low-eccentricity orbit. Wisdom (1987a, 1987b) states that these and other results indicate strong evidence that chaotic behaviour is involved in the formation of the Kirkwood gaps.

Contopoulos (1985) has shown that a transition to deterministic chaos can occur in the evolutionary dynamics of some galactic models, thereby "explaining" the high degree of structure which is present in the universe. While these models are at present quite primitive, they do indicate the significance deterministic chaos might play in the evolution of the Universe.

Hénon and Heiles (1964) seem to have initiated the study of deterministic chaos in astronomy. Their studies of the orbit of a star in a cylindrically symmetric gravitational potential reveals that, when the strength of the potential increases appropriately, the motion of the star becomes unstable. A particular discrete 2-D map (named after Hénon), now figures frequently in the literature on chaos (Campbell *et al.* 1985).

Chapter 4

History of Dynamical Systems

This Chapter is a brief history of dynamical systems. It has been compiled from the literature published in leading journals and books on the subject. Its purpose is to establish connections between the various researchers, their discoveries, and dates of their discoveries. It is hoped that the reader will thereby acquire a clearer and more meaningful perspective of the literature on deterministic chaos. More detailed information can be obtained from the many cited references.

From the earliest historical times up to the era of general relativity, the dynamical system which has most excited human interest has been the cosmos. The first essentially non-mystical geometric picture of the cosmos, due to Eudoxus and Ptolemy, consisted of the heavenly bodies circuiting the earth in fixed orbits consisting of concentric spheres and epicycles (*cf.* Dugas 1957). By contemplating 'heavenly' spheres the ancients took a momentous step. They began to study a natural system by means of a mathematical model, which enabled them to make theoretical calculations and predictions. The model of Eudoxus and Ptolemy was quite successful for calculating the dynamics of the solar system, but suffered from being too flexible. It was always possible to insert an extra epicycle to fix up a discrepancy with observation, but there was no guiding principles which might point the way to the systematic development of a better model.

Kepler's approach was fundamentally different (*cf.* Dugas 1957, page 108). Instead of looking only for a mathematical model that would closely fit observation, he also sought inner laws governing these observations, and developed models that would exhibit these laws. Kepler looked for regularities in observational data, and discovered the laws: planetary orbits are elliptical, the planetary year is proportional to the two thirds power of the distance from the sun, and the radius vector from the sun to a planet sweeps out equal areas in equal times (*cf.* Hirsch 1984, page 4). The third of Kepler's laws may be regarded as the first of the great 'conservation laws' on which the science of physics rests.

Galileo was the first person to observe the cosmos through a telescope. Of equal or perhaps greater significance was the experiments he performed on balls rolling down an incline (*cf.* Dugas 1957, page 129). Through these experiments he discovered laws relating time, distance and velocity. Galileo did not consider these quantities to represent any species of mysterious qualities. On the contrary, he took the enormously significant step of treating them as variables to be externally measured and mathematically computed. Galileo founded the first field of modern science,

dynamical systems, as a branch of mathematical physics (Galileo 1638).

Newton and Leibniz each discovered calculus, but Newton did more. He discovered the law of universal gravitation, and defined it in precise mathematical form from which the cosmic dynamics could be derived (Newton 1726). Kepler's and Galileo's laws became rigorous mathematical consequences. Newtonian mechanics was further developed by Euler, Lagrange, Laplace and others. According to Boyer (1968, page 481), Euler did for the calculus of Newton and Leibniz what Euclid had done for geometry. Euler took calculus and made it part of a more general branch of mathematics known as analysis, which is the study of infinite processes. Lagrange attempted a rigorisation of the foundations of analysis, causing a deep influence on later mathematical research (Eves 1969, page 362). Lagrange and Euler are perhaps best known for their founding work in the calculus of variations. According to Eves (1969, page 362) Euler wrote with a profusion of detail and a free employment of intuition, whereas Lagrange wrote concisely and with attempted rigour. Lagrange was modern in style and has been characterised as the first true analyst.

As a result of their work on the calculus of variations Lagrange, Euler and Maupertus established 'least action' as a principle of mathematical physics. Lagrange formulated the laws of dynamics using this principle. Hamilton and Jacobi extended and generalised Lagrange's work (Hirsch 1984). This resulted in the provision of a broad principle from which the laws of dynamics could be derived. According to Boyer (1968, page 624), from the standpoint of history, the work of Hamilton and Jacobi is significant because it prompted further research, not only in the calculus of variations, but also on systems of ordinary and partial differential equations.

Laplace's most significant work was done in the fields of celestial mechanics, probability theory, differential equations and geodesy (*cf.* Eves 1969, page 372). Laplace published two monumental works which embraced all previous discoveries in their fields along with Laplace's own contributions. Determinacy was unquestioned by Laplace, and according to Ford (1983), stated, "We ought then to regard the present state of the Universe as the effect of its preceding state and as the cause of its succeeding state." The study of differential equations has spawned whole fields of abstract mathematics. Poincaré who contributed much to differential equations also invented algebraic topology (Poincaré 1899). Lie (Hirsch 1984) invented the system of groups named after him for the same purpose. Fourier (1822) invented his series to study the heat equation and Cantor (Hirsch 1984) was led to topology and set theory by convergence problems in Fourier series.

Poincaré (1899) introduced the qualitative study of solution curves, as opposed to the quantitative study. Earlier analysts had studied individual solutions in isolation. However, Poincaré systematically studied the mutual relations between solutions, and established the idea of generic behaviour, through the concepts of stability, periodicity and recurrence (Hirsch 1984). There are two kinds of stability, stability with respect to perturbation of initial conditions, and stability with respect to perturbation of the equations themselves. Since it is not possible to know the initial conditions or equations exactly, it is necessary to seek results that are generic in the sense that they hold for most systems and for most initial conditions. Andronov and Pontryagin (Chillingworth 1976, page 229) introduced the notion structural stability (refer to §1.1) to characterise the stability of systems with respect to perturbations of their equations.

Poincaré (1899) also found one of the primal sources of chaotic dynamics, a phenomenon he termed 'transverse homoclinic points' (refer to §1.1.10). He showed that in any neighbourhood of a transverse homoclinic point there exists many different kinds of behaviour, making the system exhibit extreme instability with respect to initial conditions. The possibility of classifying or even describing the dynamics seemed to Poincaré to be remote. Poincaré's study of transverse homoclinic points is very remarkable because he did not have any examples or even a proof that one existed, but he was convinced that such points did exist and were very common; his insight proved to be correct (Hirsch 1984, page 21). The phenomenon of transverse homoclinic points was studied later by Birkhoff (1927) who proved that for three-dimensional systems every neighbourhood of such a point contains infinitely many periodic orbits. Smale (1967) extended this to n dimensions and proved that transverse homoclinic points represent a structurally stable phenomenon.

Cartwright and Littlewood (1945) showed that a rather simple differential equation could behave in a complicated way. They studied the driven van der Pol oscillator (a vacuum tube oscillator which is synchronised to an external source) (van der Pol 1934). Their discovery was that as certain parameters of the circuit were altered, the response changed through regimes of phase locking (onto various multiples of the driving signal) with apparently random motion between the regimes. Van der Pol and van der Mark (1927) discovered similar behaviour in a different oscillator, but failed to recognise its significance. Levinson (1949) (*cf.* Kloeden and Mees 1985, page 735) studied a modified version of the driven van der Pol oscillator. Smale (1967) built on this work and showed that a Poincaré map (refer to §1.1.4) for the modified system was a horseshoe shaped diffeomorphism (refer to §2.3) containing a transverse homoclinic point. Smale (1967) succeeded in explaining the general occurrence of complicated behaviour in the presence of transverse homoclinic points. However, it was not until recently that Levi (1981) (*cf.* Kloeden and Mees 1985, page 735) showed that Smale's horseshoe map was a correct description of Levinson's modified van der Pol equation. According to Casti (1982) it is generally the case that the step from a general model of the geometry to a proof that it represents the dynamics of a specific example is highly non-trivial.

Smale, Anosov and later Newhouse (*cf.* Guckenheimer and Holmes 1983; Chua *et al.* 1983b) unearthed examples of systems revealing a wealth of structurally stable dynamic behaviour in higher dimensions that do not exist in two-dimensional flows of the kind that Poincaré and Birkhoff considered. It is interesting that, in all such examples, the key points are the intersections of stable and unstable manifolds (refer to §1.1.2), demonstrating the significance of Poincaré's original insight. Smale's and Newhouse's examples also contained nontransverse intersections of stable and unstable manifolds. Such systems cannot be structurally stable. Smale and Newhouse showed that such systems form a dense set in the space of all systems, and that the set of structurally stable flows in higher dimensions do not form a dense set! Smale began to characterise the structurally stable set. This resulted in Smale defining a property called Axiom A (refer to §1.1.9), which has been the basis for almost all further work on structural stability.

Lorenz (1963) published a numerical solution to a simplified model of thermal convection, a model which now carries his name. He discovered that in this completely deterministic system of three ordinary differential equations, all non-periodic solutions were bounded but unstable (i.e. they underwent irregular fluctuations with-

out any element of randomness introduced from the outside). Lorenz's work was published in meteorology journals and went unnoticed by those physicists who could appreciated its significance until May (1974) brought it to general attention. Ruelle and Takens (1971) studied the onset of turbulence in a contained fluid heated from beneath and observed that instabilities and turbulence often occurred after three or four bifurcations. They suggested turbulence was due to a new type of attractor which they termed strange. They coined the term 'strange attractor', though they were unaware that the Lorenz model was the first example having strange attractors. The strange attractor represented a new possible mechanism for turbulence, which hitherto had been thought to be due to an infinite sequence of instabilities, a theory suggested by Landau (1959). A fundamental characteristic of motion on a strange attractor is that the asymptotic behaviour is sensitively dependent on the initial conditions in the sense that nearby trajectories move apart exponentially quickly. This is commonly called the 'butterfly effect' or the 'blunderbuss effect' (Mees 1981).

Li and Yorke (1975) introduced the word 'chaos' into the mathematical literature to denote the apparently random behaviour of some mappings, although the use of the term chaos in physics dates back to Boltzmann in a different context. Li and Yorke showed that if a one-dimensional map contains a period three trajectory then this implies that chaotic trajectories also exist. This remarkable result was a rediscovery of results by Sharkovsky who had already investigated one-dimensional maps during the 1960s, the theorem is now generally known as Sharkovsky's theorem (Kloeden and Mees 1985, page 705). Sharkovsky's results however, remained unknown in the West until Li and Yorke came upon them independently. Additional results of Sharkovsky and more recent research have been combined to give a list of equivalent conditions for chaotic behaviour in one dimension (Kloeden and Mees 1985, page 705).

Results for one-dimensional maps such as Sharkovsky's theorem do not generally remain valid in higher dimensions. An extensive analysis of recursive equations on the complex plane, which can be considered as special two-dimensional maps, was undertaken by Fatou and Julia (Peitgen and Richter 1986). These results lay for the most part dormant until the late seventies, when computer graphics enabled Mandelbrot (1977) and others to explore the rich dynamical behaviour of these maps. It is perhaps remarkable (or perhaps because of it!) that Fatou and Julia achieved as much as they did without the aid of a computer. For higher dimensional real valued recursive equations, the mathematical theory has developed in two distinct directions, depending on if the mapping is invertible (usually a diffeomorphism), or non-invertible. The study of diffeomorphisms comes under differentiable dynamics and are obtained through the Poincaré map (refer to §1.1.4). In many genuinely discrete models in the biosciences the mappings are not one-to-one, and hence are not invertible. In a generalisation of Sharkovsky's theorem, Marotto (1978) (*cf.* Kloeden and Mees 1985, page 735) shows that the presence of a special type of homoclinic point implies the existence of chaos. This area of research seems to be attracting much attention and is still rather fluid.

More detailed information has been obtained for a restricted class of recursive equation. Allwright (1978) observed that the sign of the Schwarzian derivative S (which is a measure of curvature) of a mapping plays a significant role. Allwright showed that in maps with a negative S, local stability implies global stability, which means there can be at most one stable equilibrium point or trajectory. Singer (1978), Guckenheimer (1979) and Preston (1983) provided a comprehensive analysis of such

recursive equations and their possible modes of behaviour.

Feigenbaum (1978, 1979), discovered universal properties in one-dimensional maps. Such properties were first noted by Myrbery (1958) and independently by Grossmann and Thomae (1977) and Coullett and Tresser (1978) (*cf.* Hao 1984). However, Feigenbaum emphasised the universality of such properties. At the time the physical implications of Feigenbaum's discovery was rather unclear. During the following years however, numerical and theoretical studies established this universality in a number of models. In 1980, the universality was verified in an actual turbulence experiment (Hao 1984; Pei *et al.* 1986). The work of Feigenbaum triggered an upsurge of research interest among physicists. According to Nicolis (1986, page 902), "there is no doubt that it (Feigenbaum's discovery) constitutes one of the major scientific breakthroughs of the last decade".

Sudden changes in strange attractors can occur as a system parameter is changed. Yorke (Grebogi *et al.* 1983) coined the term 'crises' to describe this behaviour. A crises is defined as a collision between a strange attractor and a coexisting unstable fixed point or periodic trajectory. Crises are involved with sudden changes in size of strange attractors, the sudden appearance of strange attractors (a new possible scenario preceding chaos; refer to §1.1.11) and the sudden destruction of strange attractors, all of which may have important practical implications.

The topological (i.e. geometric) properties of strange attractors are intricate and complicated. Usually (but not always) strange attractors have a non-integer dimension (refer to §1.1.12). Several definitions of non-integer dimension have been defined (*cf.* Farmer 1982; Grassberger and Procacia 1983; Farmer *et al.* 1983). An interesting conjecture by Kaplan and Yorke (Farmer *et al.* 1983) established a relationship between the dimension of a strange attractor and its Lyapunov exponents (refer to §1.1.12). Many experimental results have been since published demonstrating the conjecture is approximately correct (Russel *et al.* 1980). It is amazing that seemingly static properties like dimension can be related to dynamical properties like the Lyapunov exponents.

A concept which is useful in the topological study of strange attractors is the fractal; promoted initially by Mandelbrot (1977, 1987). Mandelbrot questioned why it was not possible to describe many natural objects (for example the shape of a coastline), using standard geometric shapes like circles, squares, etc. A fractal is a shape that is scale invariant (i.e. there exists an infinite number of magnifications (or scalings) of the fractal under which it looks similar or exactly the same). This implies the length of a fractal depends on the length of the measuring ruler (i.e. the smaller the ruler used the longer the measured length of the fractal). A fractal is of infinite length, but has a non-integer spatial dimension (according to definitions of dimension that Mandelbrot and others have developed). Many natural objects are fractal like (i.e. are only scale invariant for ten or so magnifications) (*cf.* Peitgen and Richter 1986; West and Goldberger 1987). The concept of a fractal has been found useful for the determination of the non-integer dimension of strange attractors. A number of introductory articles on fractals have been collected into an informative book by Peitgen and Richter (1986).

Results for differential equations are not on the whole as well-developed as those for recursive equations. When they can be reduced to diffeomorphisms via the Poincaré map, the results from recursive equations can of course be used. How-

ever, a Poincaré map does not always exist, for example at an equilibrium point. A theorem due to Shilnikov (Chua 1987, page 1064) gives conditions under which a three-dimensional flow with a certain kind of homoclinic trajectory to an equilibrium point can become chaotic. A theorem due to Melnikov (Lichtenberg and Lieberman 1983, page 240) give conditions under which a system possesses a transverse homoclinic point by showing it is a small perturbation of another system which is already known to have one. Both Shilnikov's and Melnikov's theorems rely on perturbation analysis. Perturbation methods generally do not rely on generic properties of equations and general results, but on perturbations of solutions of specific problems. This indicates the value of studying specific solutions. Arnold (1973), Smale (1967) and others make similar observations, however, this observation by Casti (1982, page 301) is colourful; "All current indications point toward the conclusion that seeking a completely general theory of nonlinear systems is somewhat akin to the search for the Holy Grail: a relatively harmless activity full of many pleasant surprises and mild disappointments, but ultimately unrewarding. A far more profitable path to follow is to concentrate upon special classes of nonlinear problems, usually motivated by applications, and to use the structure inherent in these classes as a guide to useful (i.e. applicable) results".

Dynamical processes with spatial variations or time delays cannot to described by ordinary differential equations on a finite-dimensional state space, but require a function space setting. Examples include the Navier-Stokes equations (*cf.* Swinney and Gollub 1978), reaction-diffusion equations in chemical kinetics such as the Zhabotinskii-Belousov reaction (*cf.* Tomita 1982), and delay differential equations arising in epidemiology and physiological control systems (*cf.* Farmer 1982; Glass *et al.* 1987). Such infinite-dimensional systems have been investigated extensively, both theoretically and numerically. Many exhibit seemingly chaotic behaviour for particular parameter values. For specific models the analysis required to verify the presence of a strange attractor or homoclinic trajectory is definitely non-trivial, so even more than in the case of ordinary differential equations much of what has been discovered is of a numerical or speculative nature rather than rigorously proven (Casti 1982).

Differential equations form the mathematical basis for most models of natural systems. An alternative and a complementary basis for mathematical modelling of nature is a concept called *cellular automata.* According to Wolfram (Farmer *et al.* 1984, page vii) cellular automata have five fundamental defining characteristics: 1) they consist of a discrete lattice of sites, 2) they evolve in discrete time steps, 3) each site takes on a finite set of possible values, 4) the value of each site evolves according to the same deterministic rule, and 5) the rules for the evolution of a state depend only on a local neighbourhood of sites around it. Cellular automata are constructed from many identical components, each simple, but together capable of complicated behaviour. Cellular automata can provide discrete models for homogeneous systems with local interactions or they can also be considered as idealizations of partial differential equations (in which time and space are assumed discrete, and the dependent variables take on a finite set of possible values). Cellular automata may serve as suitable models for a wide variety of systems (e.g. the growth of dendritic crystals (such as snowflakes), the patterns of pigmentation found on many mollusc shells, etc.). An empirical study by Wolfram (1984) suggests that patterns generated by cellular automata (after evolving from simple seeds), can take on one of four qualitative forms: 1) disappear with time, 2) evolve to a fixed finite size, 3) grow indefinitely at a fixed

speed, 4) grow and contract irregularly (reminiscent of chaos).

Through the 1950s two competing views on the behaviour of integrable conservative systems (refer to §1.2) perturbed by small nonlinear terms existed (Walker and Ford 1969). One view was that the nonlinear perturbation only slightly affected the dynamics of the unperturbed system, while the second view stated that it had a profound effect on the unperturbed dynamics. Kolmogorov outlined a theorem which has provided a cornerstone for linking these two divergent views (Walker and Ford 1969). Kolmogorov did not present a proof of his theorem, but the missing proof was supplied a decade later by Arnold and independently by Moser (Lichtenberg and Lieberman 1983). This theorem is consequently known as the Kolmogorov-Arnold-Moser (KAM) theorem (refer to §1.2.2). KAM theory states that for most initial conditions, a weakly perturbed conservative system generates non-ergodic motion, while for a small set of initial conditions the notion is seemingly stochastic. Mathematical attempts to characterize the stochasticity as ergodic, mixing, etc., have not generally been successful. Sinai (1963) has proved these properties for some examples. The seemingly stochastic motion is suggestive of a fundamental source for statistical mechanics, and has increasingly attracted researchers from that field. Chirikov (1960) investigated the transition between regular and stochastic dynamics, from which a criterion for this transition, called 'resonance overlay', was established (*cf.* Lichtenberg and Liebermann 1983).

There is a marked difference between the stochasticity encountered in conservative systems with two degrees of freedom (refer to §1.1) and that encountered in systems with more than two degrees. Arnold (1963) showed that for systems with more than two degrees of freedom the stochastic layers are interconnected to form a web, and leads to a global diffusion. This mechanism is commonly called Arnold diffusion (Chua *et al.* 1983, page 698).

Most conservative systems are non-integrable and have a state space with both regular and chaotic motions tangled up together in a complicated way. How trajectories wander or diffuse within the chaotic regions is important in assessing the stability of these systems. It has been found that tori having a cantor set structure (refer to §2.1) separate some chaotic regions; these separatrixes (refer to §1.1.2) are termed 'cantori', a concatenation of cantor and tori. The existence of cantori was first suggested by Percival (*cf.* MacKay *et al.* 1984). Cantori provide partial barriers to confine trajectory diffusion, structures termed 'turnstiles' provide a means for escape (MacKay *et al.* 1984).

Another related area using the qualitative approach initiated by Poincaré is Catastrophe theory (*cf.* Zeeman 1976; Golubitsky 1978; Poston and Stewart 1978; Gilmore 1981; Stewart 1981), which is concerned with the qualitative behaviour of a function near degenerate points (refer to §1.1.1). The resulting geometry can be used to describe phenomenon, particularly those in which gradually changing forces produce sudden effects. Morse and Whitney (*cf.* Woodcock and Davis 1978) established a canonical form for the structure of a function near a non-degenerate point, forms that are sometimes known as Morse critical points (Gilmore 1981). Thom (*cf.* Woodcock and Davis 1978) extended these results by establishing canonical forms for degenerate points. Thom called these canonical forms *catastrophes.* There are seven catastrophes that can occur in five dimensions or less, these are known as the elementary catastrophes. Notable contributions for the proof of Thoms theories also

came from Arnold, Malgrange and Mather (*cf.* Woodcock and Davies 1978). The significance of catastrophe theory is that dynamical motion on a catastrophe can describe in a qualitative way phenomena that have sudden events (e.g. crash of a share market, the beating of a heart, civil unrest, embryology etc.). Zeeman (1976, 1977), Poston (1978), Stewart (1981) and many others promoted catastrophe theory for describing such phenomena. This however, is also where much of the initial controversy associated with catastrophe theory stems. Zahler and Sussmann (1977, page 763) actively criticized the application of catastrophe theory stating that "catastrophe theory is one of many attempts that have been made to deduce the world by thought alone ...an appealing dream for mathematicians, but a dream that cannot come true". According to Woodcock and Davis (1978) the argument against catastrophe theory was more broad based consisting of four basic criticisms: 1) the theories foundations (Guckenheimer 1973), 2) the assumptions needed to apply it, 3) details of specific applications, and 4) attitudes and style. Today most of the controversy has subsided. The theory has been established on a more concrete foundation, and it has been found useful in some specific applications (Stewart 1981).

A review by May (1976) called attention to the complicated dynamics including period doubling and chaos in some very simple populations models. This review seems to have become a classic and did much to raise the awareness of chaos to a general audience. Ford (*cf.* Flaschka and Chirikov 1988) has been a key organiser in the spread of information on nonlinear dynamics to an interdisciplinary audience. He organised an influential conference on deterministic chaos in 1977 and helped to found Physica D in 1980. This conference seemed to mark the beginning of age for deterministic chaos (Flaschka and Chirikov 1988). Many researchers have concentrated on studying specific problems, from the earliest known problems like Lorenz's (1963) equations (Sparrow 1982) and Hénon's (1976) map, to electrical circuits (Chua 1987) and lasers (Harrison 1988). During the 1980s the topic of chaos has become a fruitful area of study and consequently has split into many specialist areas. It has been humorously said that the half life of the number of people researching chaos is six months (Esande 1985, page 165). Nevertheless, the increase which has been dramatic. The increase in the number of 'chaotic researchers' is primally due to deterministic chaos having been observed experimentally in many of the applied sciences.

A comprehensive history of mechanics up to the 1950's is given by Dugas (1957). A general history of mathematics and science are covered by Boyer (1968), Eves (1969), Kline (1972), Jaffe (1984). The more recent history and contributions to dynamics can be obtained from Reid (1975), Guckenheimer *et al.* (1977), Haken (1981), Casti (1982), Campbell *et al.* (1983), Cvitanovic (1984), Hao (1984), Hirsch (1984), Campbell (1985), Kloeden and Mees (1985), Moser (1986), Berry *et al.* (1987) and Flaschka *et al.* (1988). A delightful historical account of chaos as it has developed in the West is given in the semi-popular documentary by Gleick (1987), an American journalist who has caught the flavour and intellectual excitement accompanying these new insights.

Chapter 5

Generating Deterministic Noise

This Chapter discusses methods for generating quasi-random noise using one-dimensional recursive equations (hereafter called recursive loops), characterised by a single parameter (hereafter called the gain). The operation of a recursive loop is shown diagrammaticly in figure 5.1. Consider a sequence $\{x_n : n = 0, 1, \cdots, N\}$ of numbers generated by a recursive loop. When the number x_n appears on the left of the nonlinearity, the number x_{n+1} immediately appears on its right. If the gain, denoted by g_n, is a pre-selected constant (i.e. $g_n = g$ for all n), so that

$$x_{n+1} = gf(x_n) \tag{5.1}$$

then the x_n form a *constant-gain sequence.* After a preselected interval (the unit delay indicated in figure 5.1), x_{n+1} is transferred to the left of the nonlinearity, so that x_{n+2} immediately appears on its right. Consequently, the sequence is generated at instants each of which are separated by the aforesaid interval. It is convenient to normalize the nonlinearity such that $0 \leq f(x) \leq 1$ for $0 \leq x \leq 1$, and restrict x_n and g_n to the same range as x. The nonlinearity is said to be canonical when $f(x) = 4(1 - x)x$ (i.e. the logistic map; refer to §1.1.6).

The spectrum, (computed by the fast Fourier transform (FFT) implementation of the discrete Fourier transform (DFT)), of the sequence $\{x_n : n = 0, 1, \cdots, N\}$ is here denoted by $X_i(f(x), g, x_0, N)$, where the i^{th} frequency is i/N. The power

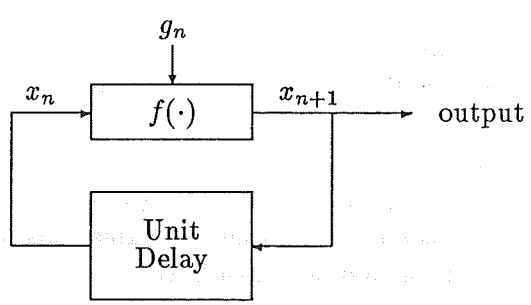


Figure 5.1: Diagrammatic representation of the recursive loop. The numbers x_{n+1} and x_n are inter-related, as indicated by (5.1), through the gain g_n which in general is different for each n, and the nonlinearity $f(\cdot)$. The output is the sequence $\{x_n : n = 0, 1, \cdots, N\}$.

spectrum PS, formed by averaging $(X_i(f(x), g, x_0, N))^2$ over M (pseudo-)randomly chosen values of x_0, the m^{th} of which is written as $x_{0,m}$, is denoted by

$$P_i(f(x), g, M, N) = (1/M) \sum_{m=1}^{M} (X_i(f(x), g, x_{0,m}, N))^2 \tag{5.2}$$

The curves in figure 5.2 show power spectra for the canonical nonlinearity for three values of g within the chaotic range (refer to §1.1.6), for $M = 20$ and $N = 1024$. The general shapes of these spectra do not change as M is increased. They merely become increasingly smooth, thereby confirming the apparently stochastic nature of the sequences.

The character of the sequence alters drastically when the gain of the loop is varied for each n, i.e.

$$x_{n+1} = g_n f(x_n) \tag{5.3}$$

where $S_g = \{g_n : n = 0, 1, \cdots, N\}$ is a sequence of gains. The x_n then form what is here called a *variable-gain sequence.* The power spectrum of the sequence $\{x_n : n = 0, 1, \cdots, N\}$ is denoted by $P_i(f(x), S_g, M, N)$, where S_g emphasises that the gain varies during the sequence.

The middle set of curves in figure 5.3 depicts the power spectra of the variable-gain sequences generated when the g_n are statistically independent and uniformly distributed from 0 to 1, and when the forms of the nonlinearities are as shown in the left hand set of curves. All three spectra have an excess low frequency character, which is to be compared with the bandpass (or even excess high frequency) character of the spectra for the constant-gain sequences shown in figure 5.2. A similar effect is obtained if the shape of the nonlinearity is kept the same but the shape of the pdf of the g_n, denoted $q_G(g)$ (refer to §2.4), is skewed. The middle set of curves in figure 5.4 depicts the power spectra of the variable-gain sequence for the $q_G(g)$ shown in the left hand set of curves in figure 5.4. As in figure 5.3, all three spectra are seen to have an excess low frequency character, taking on an increasingly 1/f form, the more the $q_G(g)$ are skewed.

Although the three nonlinearities and the three pdfs, shown in the left hand set of curves in figure 5.3 and figure 5.4 respectively, differ significantly only in their skewness, the corresponding spectra (middle set of curves in figure 5.3 and figure 5.4) exhibit significant dissimilarities. The flattening, of the dotted curves at low frequencies, and of the dashed curve for still lower frequencies, is replaced in the solid curve by a form remarkably close to 1/f. The right hand set of curves in figure 5.3 and figure 5.4 show the pdfs formed from the histograms of the variable-gain sequences. Note that these pdfs differ significantly from Gaussian.

When I discovered the simple means described above for generating a power spectrum which is very close to 1/f, I started to look for a general method for predicting the first order statistics of the x_n, given $f(x)$ and the statistics of the g_n. After extensive discussions with Professor Bates, this led to an attempt to find out what the pdf of the g_n should be, when both $f(x)$ and the statistics of the x_n are given. We also thought it worthwhile to examine whether the form of $f(x)$ could be deduced when the pdfs of both the g_n and the x_n are given. The outcomes of these investigations are reported in §5.2, §5.3, §5.4 and §5.5.

Noise is sometimes defined as 'any undesirable signal'. However, noise is not studied solely in order to reduce or eliminate it. Not infrequently, the study of noise,

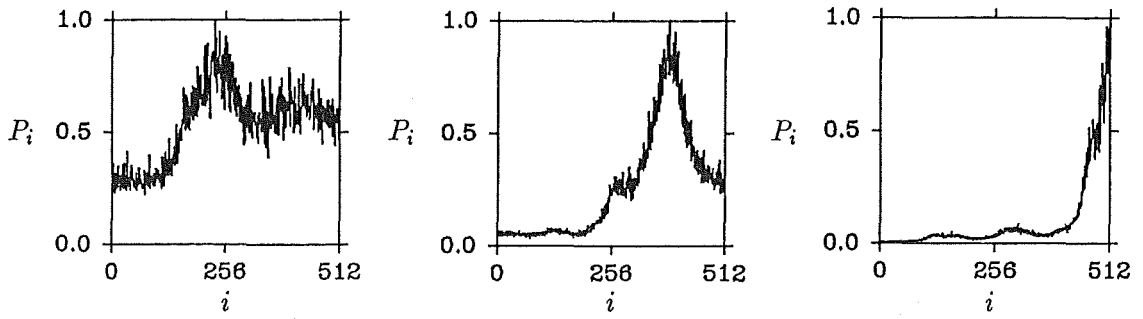


Figure 5.2: Power spectra of constant-gain sequences for canonical nonlinearity, with M=100, N=1024: g=0.9250 (left hand curve), g=0.9750 (middle curve), g=0.9975 (right hand curve). The power spectra are normalized such that their maximum values are unity. The same normalization is adopted in figure 5.3 and figure 5.4.

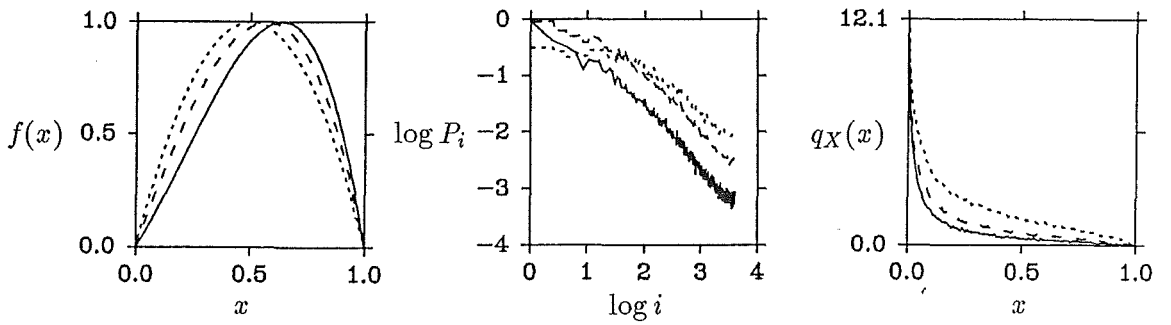


Figure 5.3: Nonlinearities and corresponding power spectra $P_i = P_i(f(x), S_g, 100, 1024)$ and pdfs of variable-gain sequences, when the g_n are uniformly and quasi-randomly distributed from 0 to 1. Left hand, middle and right hand sets of curves depict nonlinearities, logarithms (to base 10) of the power spectra and pdfs of the variable gain sequence respectively. All three power spectra are normalized such that their logarithms equal 3 when $i = 1$. —— $f(x) = h\,(x(4.5 - x)/3.5)$, $h(x) = 4x(1 - x)$, - - - - $f(x) = h\,(x(6 - x)/5)$, $h(x) = 4x(1 - x)$, ····· $f(x) = 4x(1 - x)$.

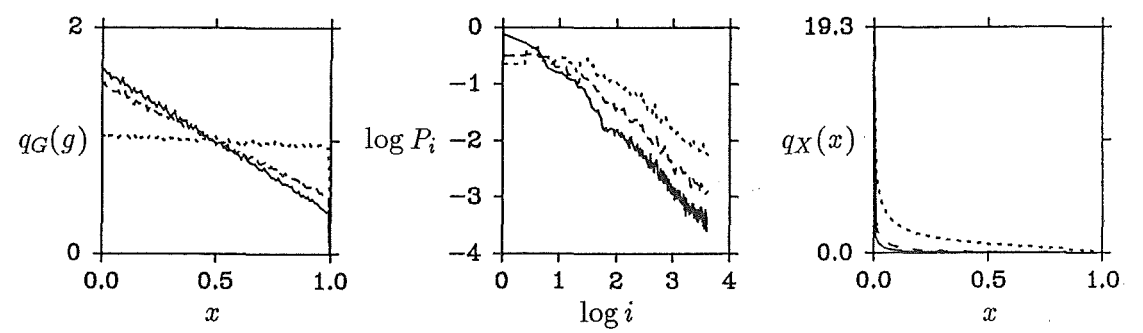


Figure 5.4: Probability density functions of g_n $q_G(g)$, and corresponding power spectra $P_i = P_i(f(x), S_g, 100, 1024)$ and pdfs of variable-gain sequences, when the $f(x)$ are canonical. Left hand, middle and right hand sets of curves depict g_n pdfs, logarithm (to base 10) of the power spectra and pdfs of the variable gain sequence respectively. All three power spectra are normalized such that their logarithms equal 3 when $i = 1$. The equation of each $q_G(g) = \alpha g + 1 - \alpha/2$. —— $\alpha = -1.2$, - - - - $\alpha = -1.0$, ····· $\alpha = 0.0$.

or the use of noise as a tool in the study of something else, has led to significant advancements (*cf.* Gupta 1975). There are many characteristics of electrical noise that make it useful. It can have: a wide bandwidth, a particular power spectrum and probability density function, a very small amplitude, be uncorrelated between nonoverlapping frequency bands, etc. Gupta (1975) considers electrical noise to have four main uses: as models of broad-band signals, for test signals, for probing physical phenomena and as a conceptual tool. §5.1 reviews the present understanding of the causes of natural noise and in particular excess low frequency noise. An interesting short history of the discovery and understanding of the many sources of natural noise is given by Johnson (1971).

The conclusions that can be drawn from the material presented in this Chapter, together with suggestions for further work, are included in Chapter 9.

5.1 Natural Noise Sources

According to Hooge *et al.* (1981, page 483) there are four main types of electronic noise: thermal, shot, generation-recombination and excess low frequency (also called Flicker or 1/f noise) (Ambrozy 1982, Chapter 4).

Thermal noise originates from the random motion of electrons arising from thermal agitation. The power spectrum of a thermal noise voltage measured across a resistor R is (Ambrozy 1982, page 101)

$$P(\mathrm{f}) = 4Rh\mathrm{f}\left[\frac{1}{2} + \frac{1}{e^{h\mathrm{f}/kT} - 1}\right] \tag{5.4}$$

where k is Boltzmann's constant ($k = 1.37\cdots\times 10^{-23}$ joules per degree), h is Planck's constant ($h = 6.62\cdots\times 10^{-34}$ joule seconds), T is the temperature in degrees Kelvin and R is resistance. For $f \ll kT/h$, (5.4) simplifies to

$$P(\mathrm{f}) = 4kTR \tag{5.5}$$

At room temperature (i.e. $T = 290°$ Kelvin) (5.5) approximates (5.4) for $\mathrm{f} \leq 10^{12}$Hz. When current flows through a resistor, the resistor ceases in general to be in thermal equilibrium, and noise additional to thermal noise may be generated (e.g. shot noise, generation-recombination noise, excess low frequency noise).

Shot noise occurs whenever charge carriers leave a cathode or pass through a potential barrier. The power spectrum for shot noise is approximately uniform up to some transition frequency, above which the power spectrum decreases. The transition frequency depends on the physical nature of the cathode or potential barrier. The uniform part of the power spectrum is characterised by (Ambrozy 1982)

$$P(\mathrm{f}) = 2qIR \tag{5.6}$$

where q is the charge of an electron, I is current, and R is the resistance of the potential barrier.

Generation-recombination noise occurs only in semiconductors. Many semiconductors have a number of charge carrier traps, a portion of which, at any one time are occupied by electrons. There is continual trapping and detrapping of electrons between the traps and the conduction band (or valence band) of the semiconductor.

Thus the number of trapped electrons, and therefore also the number of free electrons fluctuates, resulting in noise. Generation-recombination noise has a *Lorentzian power spectrum* (Ambrozy 1982):

$$P(\mathrm{f}) = \frac{4\tau}{1 + (2\pi \mathrm{f}\tau)^2} \tag{5.7}$$

where τ is the average length of time an electron is trapped. The Lorentzian power spectrum is uniform up to a frequency termed the corner frequency $f_c = 1/\tau$. Above f_c the spectral power falls at 20dB per decade of frequency. Some models for excess low frequency noise are based on generation-recombination noise.

Excess low frequency noise is a species of noise that is significant at low frequencies (less than about 1kHz). Its power spectral density seems to be inversely proportional to frequency, and because of this, it is often called '1/f noise'. No noise process can exhibit a power spectrum exactly proportional to 1/f of course because its total power would be infinite. The actual spectrum must level off at its lower end, and fall to zero faster at its upper end. According to Dutta and Horn (1981, page 499) the properties of excess low frequency noise as observed experimentally are that the exponent of f differs by less than 20% from -1 over more than 5 units of decimal logarithmic frequency scale and the pdf is Gaussian or approximately Gaussian.

The theory of most types of electrical noise is pretty well understood (*cf.* Gupta 1977; Ambrozy 1982). Theories explaining the source of electrical noise help to determine now noise can be reduced. For example, generation-recombination noise can be reduced by using cleaner preparation techniques to avoid trapping centres. On the other hand, thermal noise cannot be reduced whatever method of preparation is used, as it is fundamental and is therefore unavoidable. In this generally well understood field there is one exception: excess low frequency noise.

Temporal fluctuations with an excess low frequency power spectrum have been found in geology, music, economics, average seasonal rainfall, the rate of insulin uptake by diabetics, etc. (*cf.* Press 1978; Keshner 1982). Due to the ubiquity of excess low frequency noise, a universal cause for the phenomenon has been searched for. However to date this has been unsuccessful (Weissman 1988). The most general models for excess low frequency noise seem to poorly correlate with experiments, whereas the most successful models are specific to the system generating the noise. Even restricting attention to excess low frequency noise arising from electrical conductance, no universal models have been found. This raises the possibility of there being several types of excess low frequency noise (Gupta 1977).

When a constant current I flows through a resistor, the voltage across it is found to fluctuate with an excess low frequency power spectrum, the spectral density being proportional to I^2. Likewise, when a constant voltage V is applied across a resistor, the current flow through the resistor is found to fluctuate with an excess low frequency power spectrum, the spectral density being proportional to V^2 (*cf.* Dutta and Horn 1981; Hooge *et al.* 1981), i.e.

$$\frac{P_I(\mathrm{f})}{< I >^2} = \frac{P_V(\mathrm{f})}{< V >^2} = \frac{C_{1/\mathrm{f}}}{\mathrm{f}} \tag{5.8}$$

where $< V >$ and $< I >$ are the average values of the voltage and current fluctuations respectively, $C_{1/\mathrm{f}}$ is a proportionality constant which depends on the relative noise of the resistor, and $P_V(\mathrm{f}) \propto < V >^2 = I^2 R$ and $P_I(\mathrm{f}) \propto < I >^2 = V^2/R$ are the power

spectra of the fluctuations of the voltage and current respectively. This seems to be largely independent of the measuring conditions, such as the current or voltage level (Hooge *et al.* 1981, page 483). Equation (5.8) provides a basis for comparing measurements made in different frequency ranges and on different resistors.

At high current densities (i.e. when the resistor is forced far from thermal equilibrium), nonlinear effects become significant. The power spectral density of the voltage and current fluctuations become respectively, $P_V(\mathrm{f}) \propto < V >^{2+\alpha}$ and $P_I(f) \propto < I >^{2+\alpha}$ with $\alpha > 0$. According to Dutta and Horn (1981) there is evidence that under high current densities, a hitherto unknown noise may be generated, although evidence for it is so far tenuous.

There is evidence to indicate that fluctuations in resistance are responsible for excess low frequency noise present in voltage and current fluctuations (Weissman 1988). For small constant currents the power spectrum of the measured voltage fluctuations is proportional to the square of the current. This is consistent with resistance fluctuations being responsible for the voltage fluctuations, since $\delta V = I\delta R$, where $\delta V = V - < V >$, and $\delta R = R - < R >$. The same results are obtained regardless of whether they are measured under DC or AC conditions (Hooge *et al.* 1981). Voss and Clarke (1976) have measured a 1/f component in thermal noise from resistors which have a particularly large 1/f noise in the presence of a driving current. Their results are consistent with a resistance fluctuation, since thermal noise is characterised by $P(\mathrm{f}) = 4kTR$, and a 1/f component in thermal noise would occur if R or T fluctuated with a 1/f power spectrum.

Hooge (1969) proposes the following empirical relationship for 1/f noise arising in homogeneous metal conductors:

$$\frac{P_I(\mathrm{f})}{I^2} = \frac{P_V(\mathrm{f})}{V^2} = \frac{\alpha}{N}\frac{1}{\mathrm{f}} \tag{5.9}$$

where N is the total number of charge carriers in the conductor and α is a dimensionless constant with a value of about 2×10^{-3}. The constant $C_{1/\mathrm{f}}$ was introduced in (5.8) to normalize measurements made at different currents, voltages and frequencies. Hooge (1976) separates out two components from $C_{1/\mathrm{f}}$, by assuming that each charged carrier makes its own independent contribution to the noise (i.e. each charge carrier contributes an amount α), and the absolute noise of the conductor is divided by the number of charge carriers N (i.e. $C_{1/\mathrm{f}} = \alpha/N$). Surprisingly, α turns out to be a constant for a large (but not all) number of metal conductors. No theoretical model underlines (5.9), but the factor $1/N$ suggests that some independent fluctuations are occurring on each charge carrier, perhaps due to variations in the mobility or density of carriers. Hooge's formula (5.9) has motivated many theories based on carrier independence (*cf.* Weissman 1988). Hooge *et al.* (1981) have suggested that carrier mobility fluctuations can cause resistance fluctuations. Two 1/f models that have been extensively studied, and based on independent carrier mobility fluctuations, are the temperature fluctuation model (Voss and Clarke 1976) and Handel's (1980) quantum model.

According to Weissman (1988), any fluctuations characterising individual charge carriers cannot persist for longer than the time each charge carrier remains in the conductor. For almost all metals and semiconductors under ordinary operating conditions, the carrier transit time (and often also diffusion time) is in the microsecond to millisecond range. This would require the spectrum to flatten out below some

cutoff frequency well within the observable range. From this argument it would appear that any model based on carrier mobility or carrier number fluctuations must be wrong (Weissman 1988).

At the same time as he established (5.9), Hooge (1969) also measured the noise in gold films of varying thicknesses but otherwise identical geometry, and found that the excess low frequency noise level is proportional to the resistance, which in turn is inversely proportional to the thickness t, so that

$$P(\mathrm{f}) \propto t^{-1} \tag{5.10}$$

which is consistent with

$$P(\mathrm{f}) \propto N^{-1} \tag{5.11}$$

(*cf.* (5.9)) and is indicative of a bulk (as opposed to a surface) phenomenon. This and other evidence seems to suggest that excess low frequency noise in metals is related to a bulk effect. However, a completely different situation exists in semiconductors, where surface oxidation is important in determining the type and level of the noise. One way the surface could act as a source of noise would be through the modulation of carrier density by surface traps, as has been proposed by McWhorter (1957). The evidence seems to indicate, however, that even in semiconductors the noise is a combination of bulk and surface effects, and is not created at the surface alone.

One way to model 1/f noise is to superimpose a large number of Lorentzian spectra (i.e. $P(\mathrm{f}) \propto \mathrm{f_c}/(\mathrm{f_c^2} + \mathrm{f}^2)$) with an appropriate distribution $q_{\mathrm{f_c}}(\mathrm{f})$ of corner frequencies $\mathrm{f_c}$ (i.e. $q_{\mathrm{f_c}}(\mathrm{f}) \propto 1/\mathrm{f}$). McWhorter's (1957) model of the semiconductor oxide interface obtains such a distribution of corner frequencies. The trapping of charge carriers in traps located at different depths from the semiconductor oxide interface is the cause of the noise. Carriers become trapped for times which depend on the depth of the trap. This causes fluctuations in the numbers of charged carriers (*cf.* generation-recombination noise). The distribution of corner frequencies is determined by the distribution of distances from the traps to the interface. McWhorter's model is the most satisfactory for at least explaining part (and in some cases most) of the 1/f noise experienced in semiconductors and in particular metal oxide semiconductor (MOS) transistors (Hooge *et al.* 1981).

The physical origin of excess low frequency noise is an open question. It is not known if there is one underlining universal cause (although this now appears to be unlikely) or a number of different causes specific to individual processes (e.g. McWhorter's model for MOS transistors). However, there seems to be a consensus that excess low frequency noise results in systems which are composed of an extremely large number of subsystems, each of which operates or responds over a different time scale (relaxation time). The superposition of these subsystems result in the generation of excess low frequency noise (as in the case of the Lorentzian spectra in McWhorter's model) (*cf.* Dutta and Horn 1981, page 514; Keshner 1982; Weissman 1988). Although such an argument is appealing, there is little experimental and theoretical evidence to support it. As noted earlier, excess low frequency noise occurs in systems not in equilibrium. This raises the possibility that the same organisational principles discussed in Chapter 3 might be responsible. Webb and Gershenfield (1987) have searched for low dimensional deterministic dynamics in excess low frequency noise with no success. However, this does not rule out a deterministic cause. It seems that a new approach to explaining excess low frequency noise is in

order. An approach based on deterministic dynamics offers such an opportunity. The ease with which excess low frequency noise can be generated through deterministic dynamics has been demonstrated in the opening paragraphs of this Chapter. Such a discovery is provocative and begs further understanding. The rest of this Chapter summarises my personal attempt to uncover its mysteries.

5.2 Generating Sequences with Specified Probability Density Functions and Power Spectra

The pdf and the PS of a sequence generated by a recursive loop is characterised by the gains g_n and the nonlinear function $f(x)$. This section assesses to what extent it is possible to synthesis the gain and the nonlinear function of a recursive loop, such that the sequence generated conforms to prespecified pdf and PS.

Consider the pdf of a constant gain sequence generated by a recursive loop characterised by (5.1). If the dynamics of a recursive loop are ergodic (refer to §3.1), an invariant measure $q_X(\cdot)$ (i.e. the pdf of the generated sequence) is given by (*cf.* Collet and Eckmann 1980)

$$q_X(y) = \int_{-\infty}^{\infty} \delta(y - gf(x)) q_X(x) dx \tag{5.12}$$

where $\delta(\cdot)$ is the dirac delta function (refer to §2.3). If a sequence with a particular $q_X(\cdot)$ is required or specified, it is possible to substitute $q_X(\cdot)$ into (5.12) and solve for $gf(x)$. The $gf(x)$ thus solved for is not unique (i.e. there exists an infinite number of $gf(x)$ which satisfy (5.12) for any particular $q_X(\cdot)$). Almost all $gf(x)$ which satisfy (5.12), when substituted into (5.1) generate sequences which do not have the pdf $q_X(\cdot)$ specified (i.e. for almost all initial conditions a sequence with a different pdf is generated) (Collet and Eckmann 1980). The pdf that results for almost all initial conditions is called the *physical measure* (Eckmann and Ruelle 1985, page 626). Thus almost all the $gf(x)$ that satisfy (5.12) have a physical measure different from the pdf $q_X(\cdot)$ specified. It is not possible to determine which $gf(x)$ have dynamics with $q_X(\cdot)$ as their physical measure without actually substituting the $gf(x)$ into (5.1) and generating a sequence (which is computationally extremely expensive). In the neighbourhood of each $gf(x)$ there usually exists $gf(x)$ with similar physical measures (i.e. physical measures are usually structurally stable) (Eckmann and Ruelle 1985). Small alterations to a physical measure are made possible through (5.12), by substituting the altered physical measure into (5.12) and solving for $gf(x)$. Provided the alteration to the physical measure is sufficiently small, any $gf(x)$ solved for has the altered physical measure. The main problem, which is addressed below, is therefore to find a $gf(x)$ which has a physical measure which approximates the prespecified pdf $q_X(\cdot)$.

The only practical method for generating a constant gain sequence conforming to a prespecified pdf and PS using a single recursive loop seems to be by computer using a lookup table. The lookup table tabulates the gain, the nonlinear function, and a specification for the pdf and PS that this gain and nonlinearity generates. Someone wishing to realise a particular g and $f(x)$, enters the appropriate pdf and PS specification into the computer. The computer searches the lookup table for the closest match, from which the g and $f(x)$ are retrieved. Equation (5.12) is used to

optimise the g and $f(x)$. Such a method for generating sequences with a particular pdf and PS may have some advantages over existing methods (*cf.* Hunter and Kearney 1983; Coates *et al.* 1988). There are plenty of practical difficulties, however, such as deciding how many items are needed in the lookup table, devising how to provide concise specifications for the pdf and PS, etc.

Variable gain sequences hold greater promise, since they have an additional degree of freedom (i.e. the statistical properties of the gain sequence $\{g_n : n = 1, \cdots\}$ are adjustable). Suppose that any particular values taken on by the random variables Y and G, with g being any particular value of the latter within its range, are related by

$$y = gf(x) \tag{5.13}$$

where the value of x is specified and $f(\cdot)$ is a specified function. Since

$$q_Y(y)dy = q_G(g)dg \tag{5.14}$$

by definition (Bennet 1956), it follows that

$$q_Y(y) = \frac{1}{f(x)} q_G\left(\frac{y}{f(x)}\right) \tag{5.15}$$

Now suppose that x is any particular value taken on, within its range, by a random variable X. Then $q_Y(y)$, as defined by (5.15), is the conditional pdf of Y for a given value of x. Provided G and X are statistically independent, the actual pdf of Y is given by the right hand side of (5.15) multiplied by the pdf of X, integrated over the latter's range. If $0 \leq g \leq 1$ and $0 \leq x \leq 1$ are the ranges of G and X, respectively, then

$$q_Y(y) = \int_0^1 q_G\left(\frac{y}{f(x)}\right) q_X(x) \frac{dx}{f(x)} \tag{5.16}$$

It is convenient to identify y, g and x with x_{n+1}, g_n and x_n, respectively, as defined by (5.3) for any particular value of the integer n, suggesting that (5.16) is to be rewritten as

$$q_X(y) = \int_0^1 q_G\left(\frac{y}{f(x)}\right) q_X(x) \frac{dx}{f(x)} \tag{5.17}$$

In actual fact, G and X can only be statistically independent if

$$E(\gamma_n, \gamma_{n+m}) = 0 \quad \text{for} \quad m \neq 0, \quad \text{with} \quad \gamma_n = g_n - E(g_n) \tag{5.18}$$

where $E(\cdot)$ denotes the statistical expectation (or mean, or average). It is nevertheless found, as illustrated below, that (5.17) characterises many deterministic-chaotic processes G quite accurately.

Having specified $q_G(g)$ and the form of the nonlinearity $f(x)$, a variable-gain sequence is computed, with its pdf $q_X(x)$ being simultaneously calculated by generating a histogram from the computed values of the x_n. On substituting this histogram version of $q_X(x)$ into the right hand side of (5.17), a version of $q_X(y)$ can be generated by numerical integration. So, two versions of $q_X(x)$ are obtained, after replacing y by x in $q_X(y)$. Computational experience confirms that, when the power spectrum of the g_n, denoted by $P_i(S_g)$, is uniform (i.e. independent of i), in which case (5.18) can be expected to hold, the two versions of $q_X(x)$ differ only in the wiggles exhibited by the histogram (caused by the finite length of the sequence of the x_n). This is illustrated by the upper row of curves in figure 5.5. Even though the middle and lower

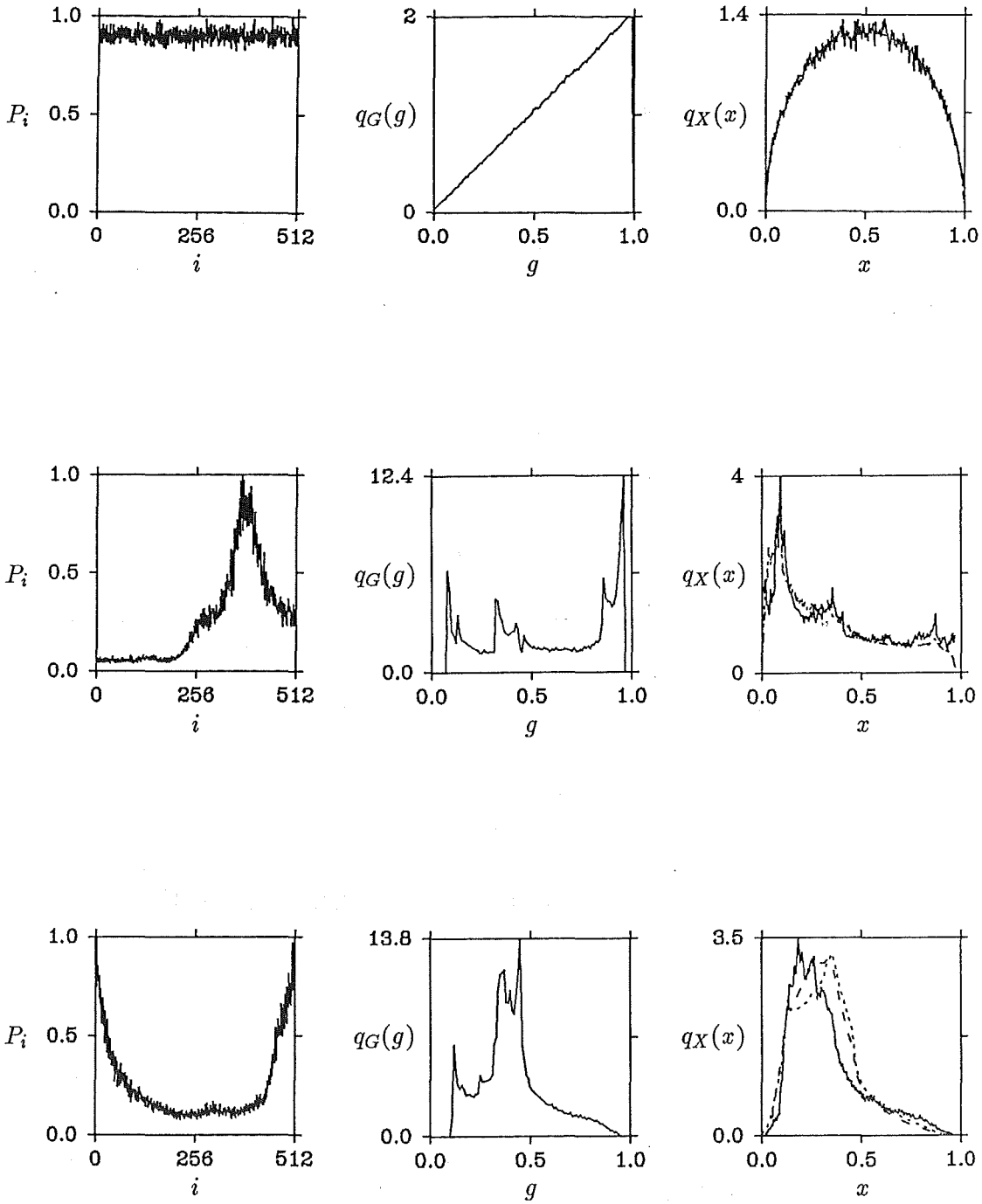


Figure 5.5: Calculated versions of $q_X(x)$ and corresponding given pdfs and power spectra (normalized such that their maximum values are unity) for canonical nonlinearity with M=100 and N=1024. Left hand column of curves: given $P_i = P_i(S_g)$; middle column of curves: given $q_G(g)$; right hand column of curves: —— given histogram versions of $q_X(x)$, - - - - versions of $q_X(x)$ obtained from (5.17) by replacing y by x in $q_X(x)$, · · · · · versions of $q_X(x)$ obtained from (5.19) after 20 iterations. In the upper and middle pairs of curves in the right hand column, only solid and dashed curves are shown because the dotted and dashed curves almost coincide.

power spectra in the left hand column of curves in figure 5.5 are far from uniform, the dashed (versions of $q_X(x)$ obtained from (5.17) by replacing y by x in $q_X(x)$) and solid (given histogram versions of $q_X(x)$) curves in the middle and lower sets of curves in the right hand column of curves in figure 5.5 are encouragingly similar, illustrating the point made in the final sentence of the previous paragraph. Note also that the dotted curves are close to the dashed curves in all three cases.

The above suggests that, if any two of $q_X(x)$, $q_G(g)$ and $f(x)$ are given, it should be possible to predict the form of the other one to a useful level of accuracy by solving the integral equation (5.17). The next section follows up this conjecture.

5.3 Inferring and Synthesizing Probability Density Functions

Given $q_G(y)$ and $f(x)$, (5.17) is solved iteratively for $q_X(x)$ by setting $q_X(x) = \lim_{m\to\infty} q_{X,m}(x)$ where

$$q_{X,m+1}(y) = \int_0^1 q_G\left(\frac{y}{f(x)}\right) q_{X,m}(x)\frac{dx}{f(x)} \tag{5.19}$$

for $m > 0$, with $q_{X,0}(x) = 1$. The dotted curves in the right hand column in figure 5.5 are the versions of $q_X(x)$ obtained from (5.19) after 20 iterations. The solid and dotted curves are encouragingly similar in all three sets of curves.

Given $q_X(x)$ and $f(x)$, (5.17) is solved numerically for $q_G(g)$ by approximating it by an array of uniformly spaced, weighted pulse functions:

$$q_G(g) \approx \sum_{k=1}^{K} A_k \lambda\left(g - (2k-1)/2k\right) \tag{5.20}$$

with the pulse function $\lambda(t)$ and a corresponding test function $\tau(t)$ chosen such that

$$\int_0^1 \lambda\left(t - (2k-1)/2k\right)\tau\left(t - (2l-1)/2k\right)dt = \delta_{kl} \tag{5.21}$$

where δ_{kl} is the Kronecker delta (zero for $k \neq l$ and unity for $k = l$). It follows that

$$A_k = \int_0^1 q_G(g)\tau\left(g - (2k-1)/2k\right)dg \tag{5.22}$$

The positive integer K is chosen large enough to allow the error implicit in the summation in (5.20) to be acceptable (this is assessed by solving (5.17) for several increasing values of K, checking that $q_G(g)$ is converging to a limiting form). Substituting (5.20) into (5.17) for $y = (2l-1)/2k$, with $l \in \{1, 2, \cdots, K\}$, gives K linear inhomogeneous algebraic equations for the K unknown A_k, which are readily evaluated by matrix inversion. Figure 5.6 shows results for $K = 50$, $\lambda(g) =$ Dirac delta function, $\tau(g) =$ rectangular function of height K and width $1/K$, and $f(x) =$ canonical nonlinearity. The right hand column of curves shows: 1) (as solid curves) versions of $q_G(g)$ giving rise to the corresponding $q_X(x)$ shown in the left hand column of curves, and 2) (as dashed curves) versions of $q_G(g)$ obtained, as described above, from (5.20) and (5.17) with $q_X(x)$ as shown in the left hand column of curves.

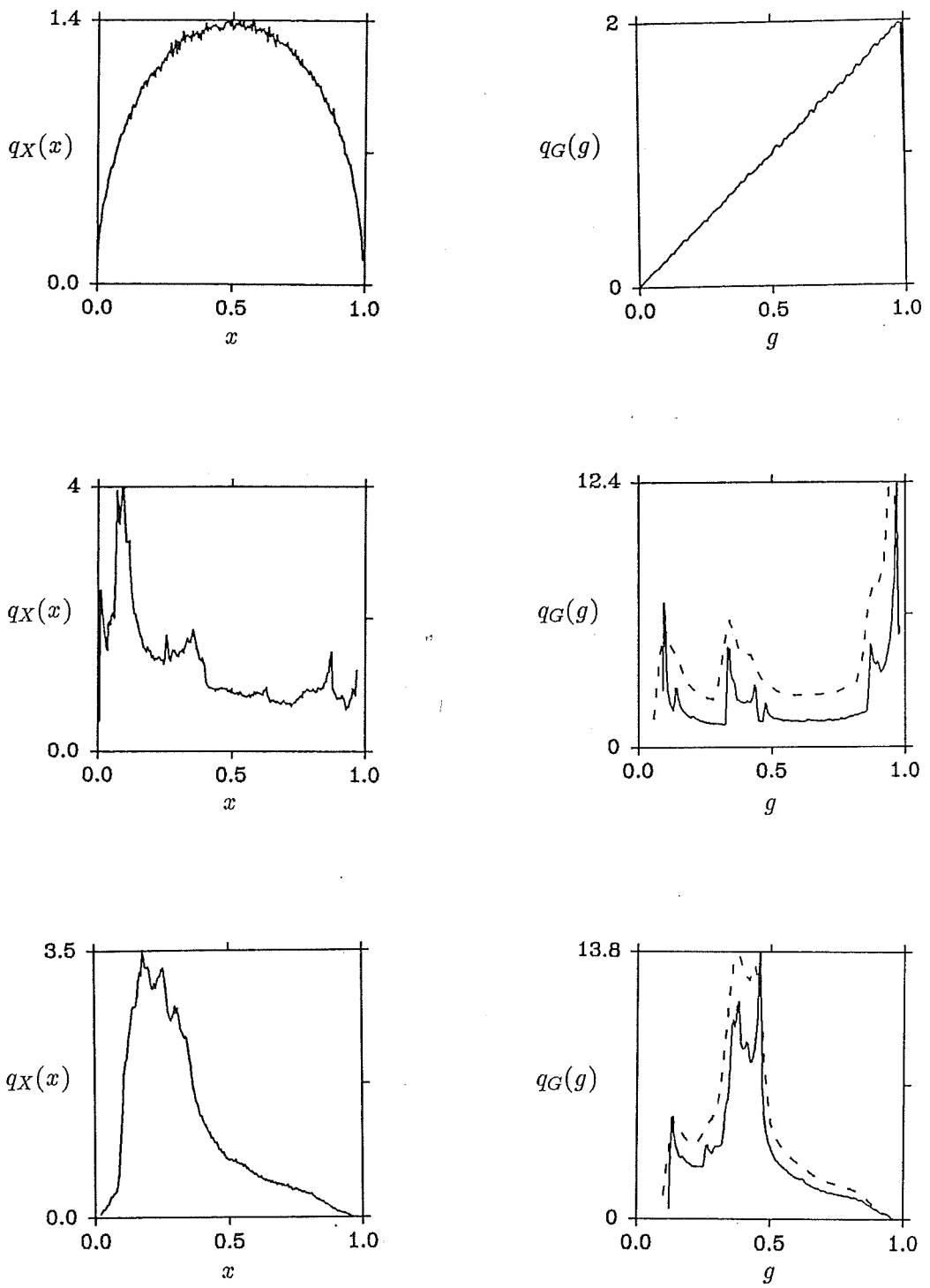


Figure 5.6: Calculated versions of $q_G(g)$ and corresponding given pdfs. Left hand column of curves: given $q_X(x)$; right hand column of curves: —— version of $q_G(g)$ shown in middle column of curves in figure 5.5, - - - - version of $q_G(g)$ calculated from (5.20) and (5.17) with $k = 50$.

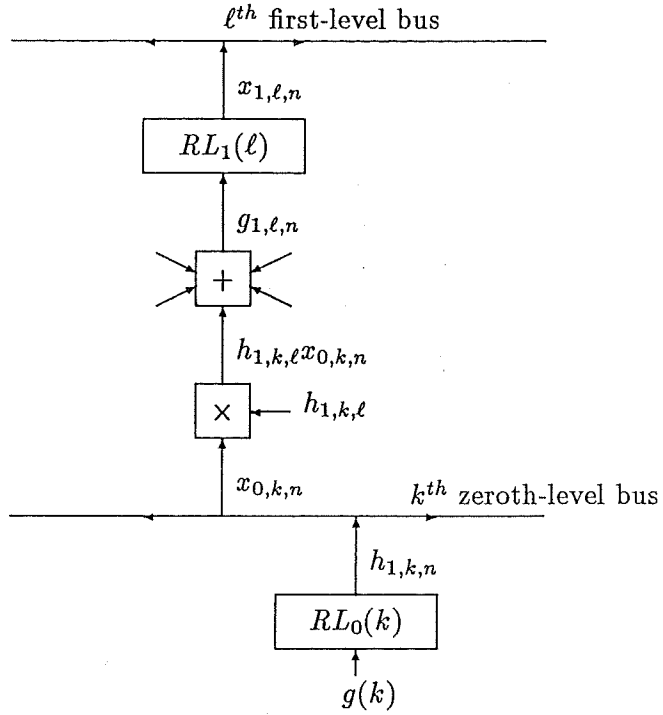


Figure 5.7: Illustration of structure of a hierarchy of recursive loops. Two such loops, and their connection through a zeroth-level bus, are shown: the k^{th} at the zeroth-level and the l^{th} at the first-level. The functions of the multipliers and summer are explained in the text, as are the meanings of the various symbols.

All three pairs of dashed and solid curves are encouragingly similar. While there are clear differences between the dashed and solid curves in the middle and lower plots of $q_G(g)$ shown in figure 5.6, the dashed curves follow the same trends as the solid curves.

Given $q_X(x)$ and $q_G(g)$, $f(x)$ is solved numerically, imposing whatever constraints are appropriate (e.g. forcing $f(x)$ and $f'(x)$ to have particular values for particular values of x). Any initial estimate, satisfying the aforesaid constraints, is chosen for $f(x)$. It is substituted, together with the given $q_X(x)$ and $q_G(g)$, into the integral in (5.17). The discrepancy between the resulting estimate of $q_X(y)$ and the given $q_X(x)$ generates a new estimate for $f(y)$ which, when y is replaced by x, is re-substituted into the integral. This iterative procedure is continued until the differences between successive estimates for $f(x)$ can be considered negligible. Computational experience suggests that, unless the aforesaid constraints on $f(x)$ are so strong as to largely fix its form, it is found that the iterations tend to lead to forms for $f(x)$ which, when substituted into either (5.1) or (5.3), do not generate chaotic sequences. It thus transpires that this third iterative approach is only useful when comparatively minor changes to $q_G(g)$ and/or $q_X(x)$ are required, such as between the forms for $q_X(x)$ shown in the right hand sets of curves in figure 5.3.

5.4 Generating Completely Deterministic Noise

The variable-gain sequences defined by (5.3) are only partly deterministic because $\{g_n : n = 0, 1, 2, \cdots, N\}$ is introduced as a stochastic sequence. It is natural to enquire whether it is possible to further develop this approach so as to generate, in completely deterministic fashion, noisy processes of arbitrary statistical character. This section discusses a generalization of the recursive loop.

A collection of recursive loops is envisaged, each typified by figure 5.1, organised hierarchically into levels. The k^{th} loop at the j^{th} level is denoted by $RL_j(k)$. The output and variable gain, both at the n^{th} instant, and the nonlinearity of $RL_j(k)$ is written as $x_{j,k,n}$, $g_{j,k,n}$ and $f_{j,k}(\cdot)$ respectively. There are L_j loops at the j^{th} level. Their outputs set the variable gains of the loops at the $(j+1)^{th}$ level. At the bottom (or zeroth) level, the gains are constant, with $g(k)$ being the gain of the k^{th} zero-level loop. A hierarchy containing J levels is said to be of order J.

There are many possible ways of structuring the hierarchy. One possibility involves specifying that, before the output of $RL_j(k)$ is fed (via what we call the k^{th} j^{th}-level bus) to $RL_{j+1}(l)$, it is weighted (i.e. multiplied) by the constant $h_{j+1,k,l}$, after which the variable gain of $RL_{j+1}(l)$ is set by summing all L_j of these weighted outputs, i.e.

$$g_{j+1,l,n} = \sum_{k=1}^{L_j} h_{j+1,k,l} x_{j,k,n} \tag{5.23}$$

figure 5.7 illustrates this. It shows typical recursive loops at the zeroth and first levels. The k^{th} zeroth-level bus feeds the output (i.e. $x_{0,k,n}$) of $RL_0(k)$ to L_1 multipliers, each of which is associated with a single first-level recursive loop (the l^{th} being shown in figure 5.7). The output of each multiplier goes to a summer, which has inputs (indicated by the five, which is only a nominal number, incoming arrows converging on the box labelled with a plus sign) emanating from all L_0 of the zeroth-level buses. The output of the summer, which is given by (5.23) with $j = 0$, sets the variable gain of $RL_1(l)$, whose output is fed to the l^{th} first-level bus. This bus feeds a typical second-level recursive loop (say the m^{th}) in the same way as the k^{th} zeroth-level bus is shown in figure 5.7 to feed $RL_1(l)$. This is expressed precisely by replacing k,l, subscript 0 and subscript 1 by l,m, subscript 1 and subscript 2 respectively. The output of a hierarchy of order J is obtained by summing the outputs of all L_J of the loops at the J^{th} level.

It has previously been demonstrated how readily excess low frequency noise can be generated by simple hierarchies, for which $L_j = 1$ for all j. Note, in particular, curves numbered 4 and 5 in the bottom sets of curves in figure 2 of Bates and Murch (1987). Presented here are some specimen pdfs and power spectra for more complicated hierarchies, not with any intent of being exhaustive (which is obviously impracticable), but merely to demonstrate the ease with which virtually arbitrary, apparently stochastic processes can be generated by deterministic chaos. Power spectra and pdfs of the outputs of hierarchies of order 2 and order 3 are shown in figure 5.8 and figure 5.9 respectively. Tables 5.1 and 5.2 list the parameters specifying these hierarchies. Note that the upper and middle power spectra in figure 5.9 each have a pronounced excess low frequency character, whereas the upper spectrum in figure 5.8 rises steeply at both low and high frequencies. The middle and lower spectra in figure 5.8 have an excess high frequency character, while the lower

Table 5.1: Hierarchy parameters for the power spectra and pdfs illustrated in figure 5.8. Each $f_{j,k}(\cdot)$ is the canonical nonlinearity. The meanings of a, b and c heading the three columns of numbers are explained in the caption to figure 5.8.

Hierarchy parameters	figure 5.8		
	a	b	c
$h_{1,1,1}$	0.925	0.75	0.5
$h_{1,2,1}$	0.0	0.175	0.5
$h_{1,1,2}$	0.0	0.2475	0.5
$h_{1,2,2}$	0.9975	0.75	0.5
$g(1)$	0.925	0.925	0.925
$g(2)$	0.9975	0.9975	0.9975

Table 5.2: Hierarchy parameters for the power spectra and pdfs illustrated in figure 5.9. Each $f_{j,k}(\cdot)$ is the canonical nonlinearity. The meanings of a, b and c heading the three columns of numbers are explained in the caption to figure 5.9.

Hierarchy parameters	figure 5.9		
	a	b	c
$h_{1,1,1}$	0.35	0.75	0.9
$h_{1,2,1}$	0.35	0.15	0.075
$h_{1,3,1}$	0.35	0.15	0.075
$h_{1,1,2}$	0.35	0.15	0.075
$h_{1,2,2}$	0.35	0.75	0.9
$h_{1,3,2}$	0.35	0.15	0.075
$h_{1,1,3}$	0.35	0.15	0.075
$h_{1,2,3}$	0.35	0.15	0.075
$h_{1,3,3}$	0.35	0.75	0.9
$h_{2,1,1}$	0.325	0.325	0.325
$h_{2,2,1}$	0.325	0.325	0.325
$h_{2,3,1}$	0.325	0.325	0.325
$h_{2,1,2}$	0.325	0.325	0.325
$h_{2,2,2}$	0.325	0.325	0.325
$h_{2,3,2}$	0.325	0.325	0.325
$h_{2,1,3}$	0.325	0.325	0.325
$h_{2,2,3}$	0.325	0.325	0.325
$h_{2,3,3}$	0.325	0.325	0.325
$g(1)$	0.95	0.95	0.95
$g(2)$	0.95	0.95	0.95
$g(3)$	0.95	0.95	0.95

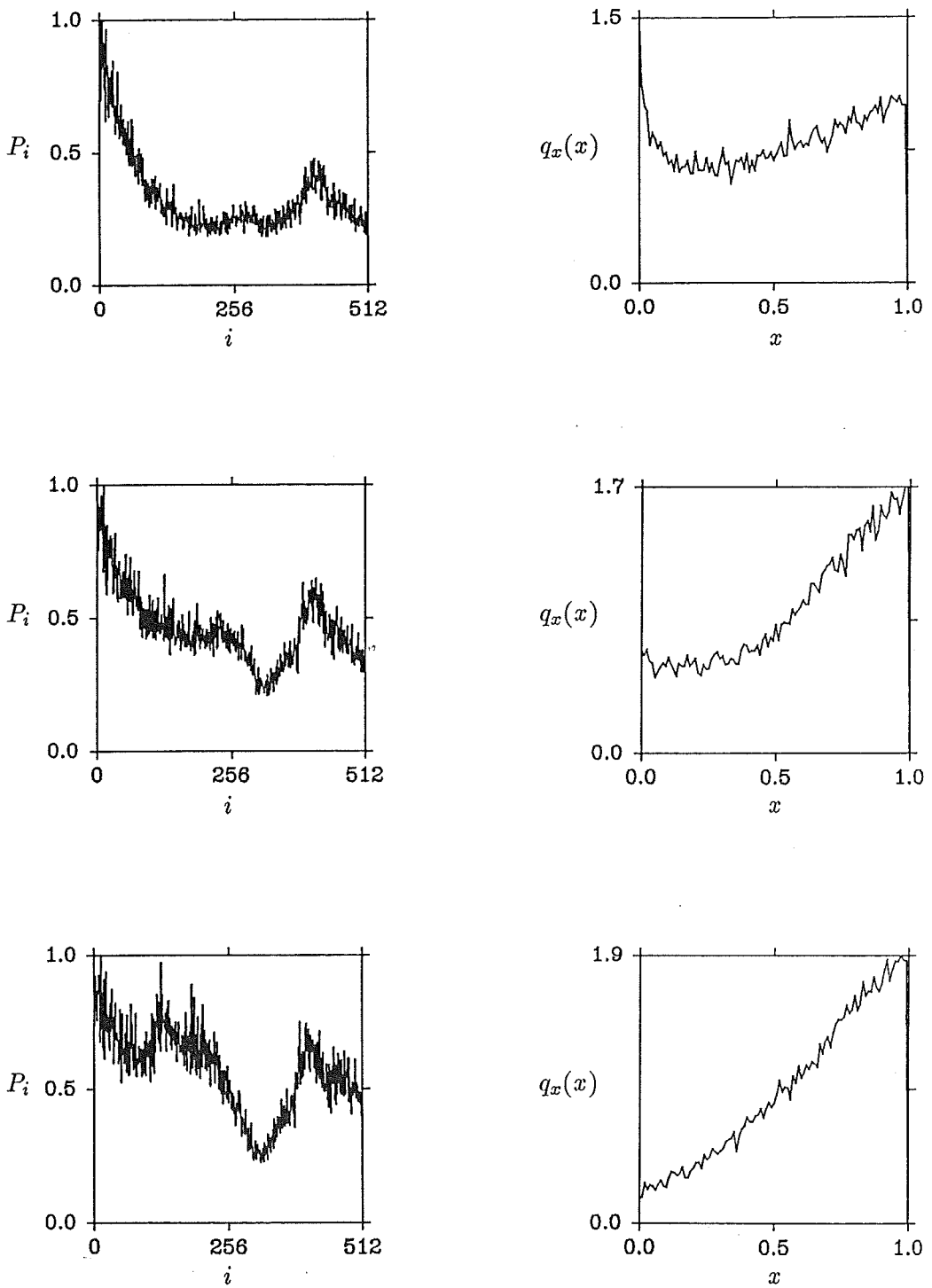


Figure 5.8: Power spectra and pdfs for three hierarchies of order 2. The parameters specifying these hierarchies are listed in table 5.1. The top, middle and bottom row of curves in the figure correspond, respectively, to the parameters listed in columns a, b and c in table 5.1.

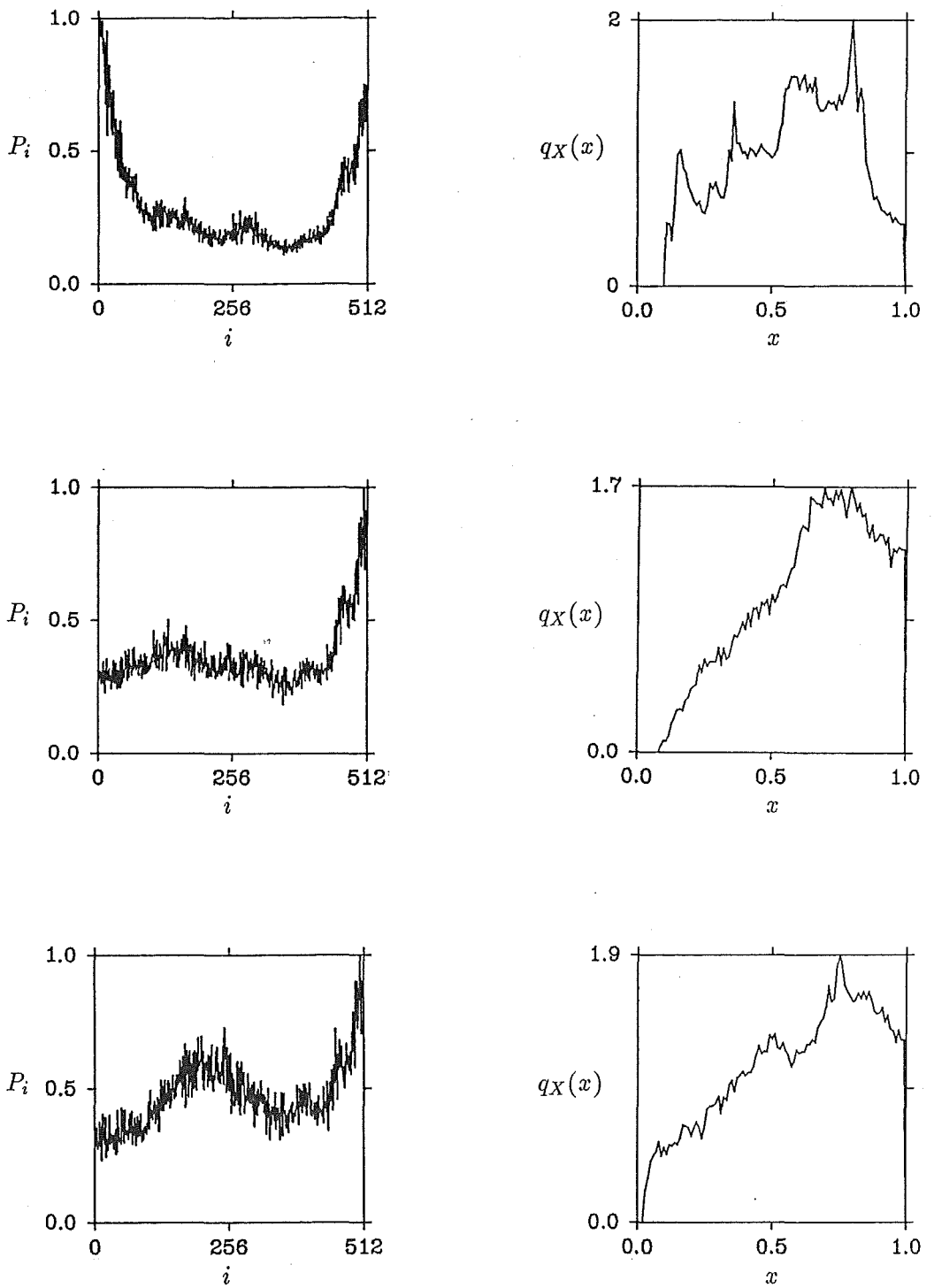


Figure 5.9: Power spectra and pdfs for three hierarchies of order 3. The parameters specifying these hierarchies are listed in table 5.2. The top, middle and bottom row of curves in the figure correspond, respectively, to the parameters listed in columns a, b and c in table 5.2.

spectrum in figure 5.9 meanders about a roughly constant level.

Some might argue that the above hierarchies are not completely deterministic because, in practice, the values of the $g(k)$ and the $h_{i,k,l}$ would have to be set by random processes, to avoid the intervention of a *deus ex machina*. This criticism is accepted, but point out that it can be satisfactorily answered by arranging feedback, from higher to lower levels, such that the $g(k)$ and $h_{j,k,l}$ are themselves members of deterministic sequences.

5.5 Further Analysis of Variable-Gain Recursive Loops

In reality a dynamical system does not operate in isolation, it is continually affected or perturbed by disturbances external to itself. The consequences of adding external noise to the trajectory of a chaotic system have been studied by several researchers (*cf.* Crutchfield and Huberman 1980; Eckmann 1981, Crutchfield *et al.* 1982, Franaszek 1984). However, this is not the only way external noise can make its presence felt. It can also perturb the system equations, and in particular the parameters characterising the system equations. This section discusses the variable-gain canonical recursive loop

$$x_{n+1} = g_n b x_n(1 - x_n) = g_n f(x_n) \tag{5.24}$$

where the g_n are statistically independent and uniformly distributed over the interval $[1, \alpha]$ (where α is a parameter in the interval $[0, 1]$), and where b is a parameter set to a value in the interval $[0, 4]$. The g_n in (5.24) can be thought of as inducing external noise into the system parameter b (where α sets the amplitude of the induced noise).

The curves in figure 5.10 depict, respectively, power spectra and variable-gain sequences. The three curves, from top to bottom in figure 5.10, are for $b = 4$ and for $\alpha = 0.2$, 0.1 and 0.0 respectively. The power spectra take on an increasingly excess low frequency character, and the variable-gain sequence becomes increasingly intermittent (i.e. time spent at small amplitudes are followed by short turbulent bursts, where the duration between bursts are seemingly unpredictable) as $\alpha \to 0$. Such intermittent sequences are known to process an excess low frequency character (*cf.* Procaccia and Schuster 1983).

Equation (5.24) has a fixed point $\bar{x}$ at zero (i.e. $\bar{x} = g_n f(\bar{x})$, for $\bar{x} = 0$). The stability of $\bar{x}$ at the n^{th} time instance is determined by g_n. The expected stability $E(g_n f'(0))$ of $\bar{x}$ is calculated by forming the long term average of the stability of $\bar{x}$ at each temporal instant n, i.e.

$$\begin{aligned} E(g_n f'(0)) &= \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} g_n f'(0) \\ &= \frac{1+\alpha}{2} b \end{aligned} \tag{5.25}$$

The fixed point $\bar{x}$ is on average stable if $| E(g_n f'(0)) | < 1$, and on average unstable if $| E(g_n f'(0)) | > 1$. For $b = 4$ and $\alpha = 0$, $\bar{x}$ is on average unstable since $| E(g_n f'(0)) | = 2$. Intermittency occurs because, whenever a trajectory is injected (by the system dynamics) near to $\bar{x}$, it spends a comparatively long time moving away (i.e. escaping) from $\bar{x}$. This results in laminar behaviour (i.e. the behaviour between turbulent

bursts; refer to figure 5.10). After the trajectory has moved sufficiently far from $\bar{x}$ the trajectory behaves erratically but is quickly injected back near to $\bar{x}$ by the system dynamics. This can be demonstrated by considering the expected value of $x_{n+1} - x_n$ for each value of x_n in the interval $[0, 1]$ (i.e. $E(x_{n+1} - x_n \mid x_n)$). This is plotted in figure 5.11(a), and shows that if x_n is close to $\bar{x}$ the trajectory is expected to move away from $\bar{x}$, but once the trajectory is far from $\bar{x}$, (i.e. $x_n >\approx 0.6$) the trajectory is expected to be injected back near $\bar{x}$. Figure 5.11(b) depicts the probability of x_{n+1} being greater than x_n for each value of x_n in the interval $[0, 1]$ (i.e. $P(x_{n+1} > x_n \mid x_n)$. Figure 5.11(b) shows that if x_n is small, x_{n+1} is more likely to be larger than x_n, while if x_n is large, x_{n+1} is more likely to be smaller than x_n. In fact, if $x_n >\approx 0.75$, it is completely certain that x_{n+1} will be smaller than x_n (i.e. has probability zero of being greater).

It is emphasised that this type of intermittency is different from the types discovered by Pomeau and Manneville (1980). The intermittency they discovered results from the presence of a marginally unstable fixed point, or the presence of an 'illusory' fixed point. The type of intermittency that results from (5.24) occurs when a fixed point is made randomly stable and unstable (but on average is marginally unstable).

Deterministic chaos is associated with sensitive dependence on initial conditions (refer to §1.1.8), and is characterised by a positive Lyapunov exponent (refer to §1.1.12). An infinitesimal perturbation (denoted at the n^{th} instance by q_n) of a trajectory grows on average exponentially with increasing n. This rate of growth is governed by the linearized equation (*cf.* Devaney 1987)

$$q_{n+1} = q_n f'(x_n) \tag{5.26}$$

The Lyapunov exponent of (5.24) (denoted $\lambda(\alpha, b)$) is a function of α and b, and is given by (*cf.* §1.1.12)

$$\begin{aligned} \lambda(\alpha, b) &= \lim_{n\to\infty} \frac{1}{n} \ln \mid \frac{q_n}{q_0} \mid \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \ln \mid g_n f'(x_i) \mid \end{aligned} \tag{5.27}$$

The Lyapunov exponent computed from (5.27) is plotted against α in figure 5.12(a) for $b = 4$ (i.e. $\lambda(\alpha, 4)$ plotted against α). The figure shows that the Lyapunov exponent is negative when $\alpha <\approx 0.6$, thus, for these values of α an infinitesimal perturbation of a trajectory q_n does not grow on average with n (i.e. (5.24) is not chaotic for $\alpha < 0.6$ and $b = 4$). For comparison, the value of $\lambda(1, b)$ for $\alpha = 1$, averaged over the interval $c < b < 4$, i.e.

$$E(\lambda(1,b)) = \frac{1}{4-c} \int_c^4 \lambda(1,b) db \quad \text{for} \quad c < b < 4 \tag{5.28}$$

is plotted against c in figure 5.12(b). It is informative to compare figure 5.12(a) with figure 5.12(b). The Lyapunov exponent plotted in figure 5.12(a) decreases with decreasing α for α down to 0.2, after which the Lyapunov exponent increases with decreasing α. This is to be contrasted with figure 5.12(b), where the Lyapunov exponent actually decreases as $c \to 0$. This demonstrates that the dynamics of (5.24) are unique, and cannot be characterised by averaging (or by other operations) the descriptors of other systems.

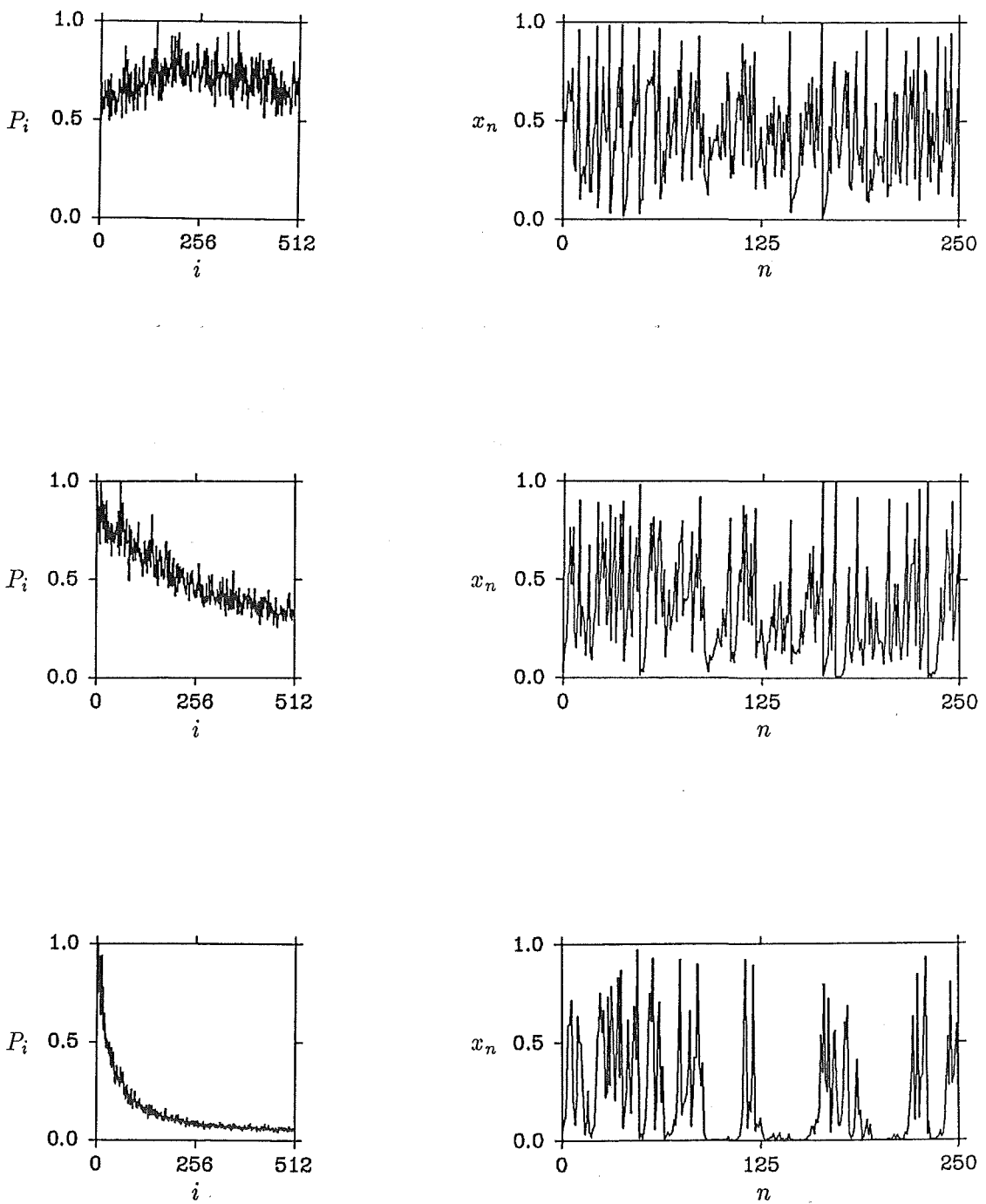


Figure 5.10: Power spectra and variable-gain sequences generated by (5.24) for b=4. Top curves $\alpha = 0.2$, middle curves $\alpha = 0.1$, bottom curves $\alpha = 0$.

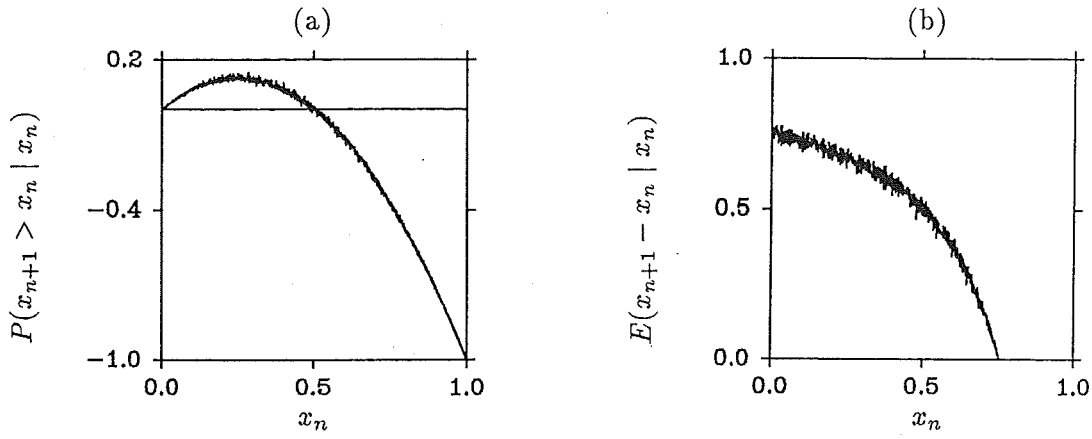


Figure 5.11: Two statistical properties of (5.24) for $\alpha = 0$ and $b = 4$. (a) The expected value of $x_{n+1} - x_n$ (i.e. $E(x_{n+1} - x_n \mid x_n)$), for each value of x_n in the interval $[0, 1]$. (b) The probability that x_{n+1} is greater than x_n, $P(x_{n+1} > x_n \mid x_n)$ for each value of x_n in the interval $[0, 1]$.

The long term average in the separation of infinitesimally close trajectories is characterised by the Lyapunov exponent $\lambda(\alpha, b)$. However, over short time scales the separation may not evolve uniformly. Figure 5.12(c) plots the logarithm of the separation in distance normalized with respect to q_0 (i.e. $\ln \mid q_n/q_0 \mid$) against temporal instant n for $\alpha = 0$ and $b = 4$. At some instants n, the separation $\ln \mid q_n/q_0 \mid$ is increasing, but on average it is decreasing. These deviations from average are considerably greater than when $\alpha = 1$ and $b = 4$. For comparison, figure 5.12(d) plots $\ln \mid q_n/q_0 \mid$ against n for $\alpha = 1$ and $b = 4$. Such deviations in the growth of $\ln \mid q_n/q_0 \mid$ are termed *nonuniformity* (*cf.* Herzel and Pompe 1987). Two time scales can be distinguished in the dynamics of (5.24), over short time scales nonuniformity dominates whereas over large time scales the growth of small perturbations is governed by the Lyapunov exponent.

To determine the extent to which a trajectory can be predicted, first introduce a covering $\{B_i : i = 1, \cdots, m\}$ of the interval $[0, 1)$ consisting of the m shorter intervals

$$\{B_i = \left[\frac{1}{m}(i-1), \frac{i}{m}\right) : i = 1, \cdots, m\} \tag{5.29}$$

Each B_i can be considered to represent the state of (5.24). The conditional entropy $H_c(k)$ (i.e. the uncertainty; *cf.* Pierce 1980) of a future state B_j after k iterations of (5.24), when the initial state B_i is known is by definition

$$H_c(k) = -\sum_i p_i \sum_j p_{j|i}(k) \ln[p_{j|i}(k)] \tag{5.30}$$

where p_i is the probability of finding the system in state B_i at any time, and $p_{j|i}(k)$ is the conditional probability of finding the system after k iterations in state B_j given that it was originally in state B_i. For $H_c(k) = 0$ the future state B_j is completely known from a measurement of B_i. On the other hand, for

$$H_c(k) = H = -\sum_i p_i \ln p_i \tag{5.31}$$

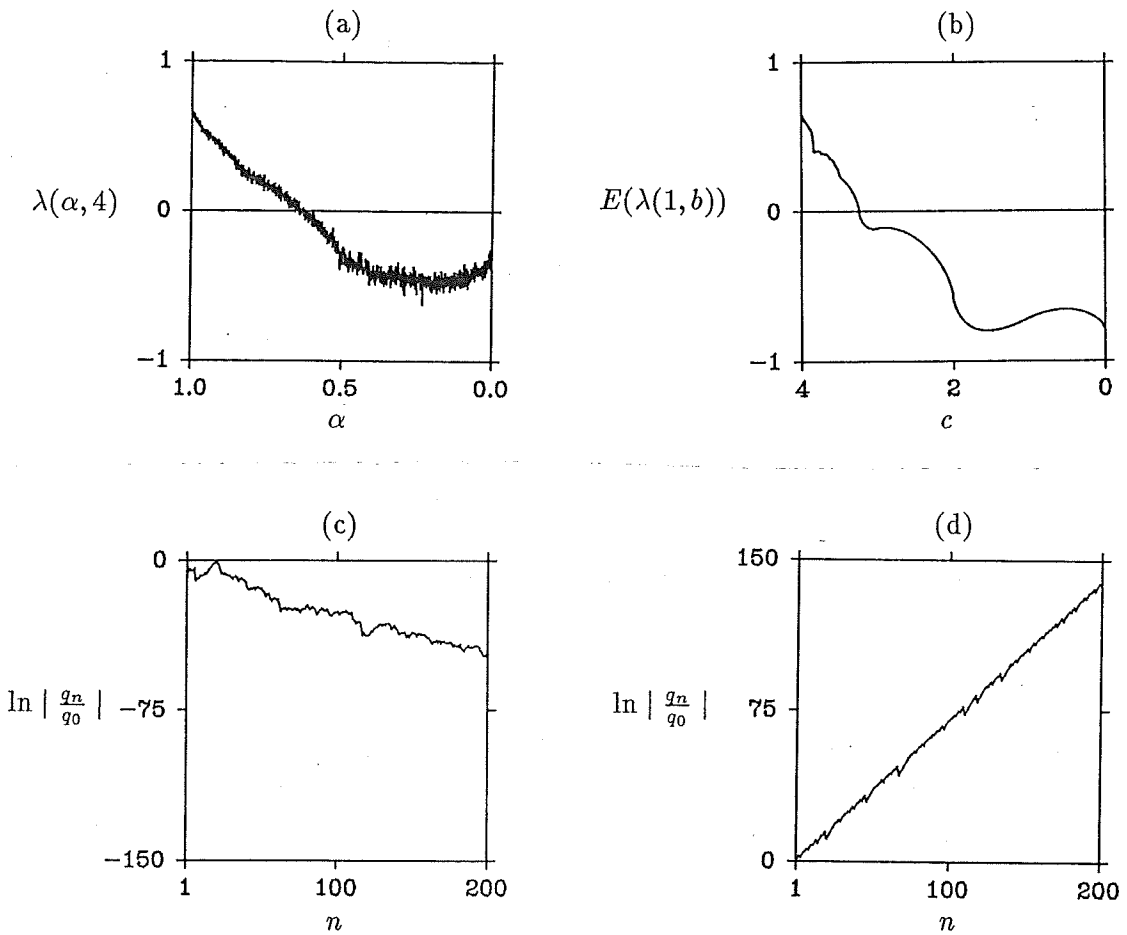


Figure 5.12: Growth of infinitesimal perturbations. **(a)** Lyapunov exponent plotted against α for $b = 4$ (i.e. plot of $\lambda(1, b)$ against b). **(b)** The Lyapunov exponent averaged over b, for b's in the interval $[c, 4]$, plotted against c. **(c)** Plot of the separation of distance $\ln \mid q_n/q_0 \mid$ at each time instance n for $\alpha = 0$ and $b = 4$. **(d)** Plot of the separation of distance $\ln \mid q_n/q_0 \mid$ at each time instance n for $\alpha = 1$ and $b = 4$.

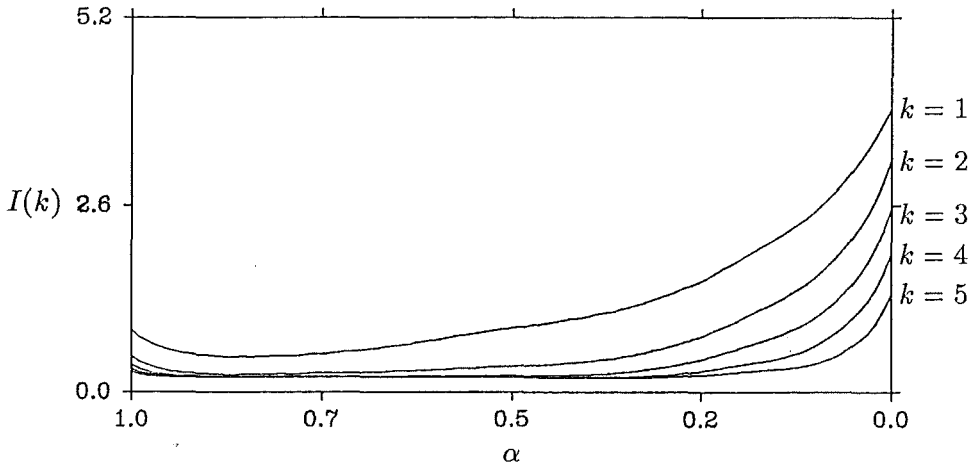


Figure 5.13: Plot of $I(k)$ against α, for k=1 to 5 and for m=100. Note that $H = 5.2$, and that similar results are obtained when $m = 20$ and when $m = 200$.

a measurement of the present state fails to provide any information about the future. The mutual information $I(k)$ defines the amount of information known about the system after k iterations of (5.24), when the original state of the system is known (*cf.* Pierce 1980), where

$$I(k) = H - H_c(k) \tag{5.32}$$

Figure 5.13 shows $I(k)$ plotted against α for k=1 to 5 and m=100.

Figure 5.13 shows that mutual information decreases with increasing noise amplitude (i.e. $\alpha \rightarrow 0$) for α down to ≈ 0.9, after which the mutual information increases with increasing noise amplitude. The mutual information (i.e. the degree with which the future can be predicted) increases in spite of increasing noise amplitude and increasing Lyapunov exponents (*cf.* figure 5.12(a)) both of which generally reduce the predictability in deterministic systems. This improvement in predictability with increasing noise amplitude (for $\alpha \rightarrow 0$) is here termed *noise induced predictability*. It has been conjectured that nonuniformity is a necessary condition for the occurrence of noise induced predictability (*cf.* Herzel and Pompe 1987). These results provide further evidence in support of this conjecture.

The significance of these results is that for some chaotic dynamical systems increasing the amplitude of additive noise actually increases the predictability of the system behaviour (i.e. reduces future uncertainty). There are however, many questions which remain open, such as: does noise-induced predictability occur in the majority of systems? To what extent does noise-induced predictability increase predictability? How should the noise be introduced? Although noise is already sometimes introduced into control systems (to reduce stiction and other undesirable aspects of mechanical activators), additive noise has hitherto not been considered as a means of improving the predictability of system behaviour. Noise-induced predictability may be of technological importance since it might allow systems and processes which are at present unpredictable (e.g. system reliability, weather), to be made (more) predictable and useful.

Chapter 6

Sinusoidal Oscillator Noise

The recursive loop depicted in figure 6.1 is a generalisation of the simple recursive loops (capable of generating chaotic discrete signals) studied in Chapter 5. Since each of the K delays in figure 6.1 can be different, the recursive loop is said to possess memory because the input to the nonlinearity $f(\cdot)$ depends in general upon K previous outputs. The recursive loop with memory can be regarded as a primitive discrete time oscillator operating at baseband (i.e. it oscillates at zero frequency!), where the K parallel attenuating/delay paths are analogous to a narrow band filter. Analysis of the loop is eased (without any apparent loss of significant content) by taking the individual attenuations to be integer powers of a minimum attenuation α, and the individual delays to be multiples of a minimum delay τ_0, i.e.

$$\alpha_k = \alpha^k \text{ and } \tau_k = k\tau_0, \text{ for } k = 1, \cdots, K \tag{6.1}$$

If the subscript n, on the symbols x_n and y_n shown in figure 6.1, is understood to increase by unity after each interval of duration τ_0, inspection of figure 6.1 indicates that

$$x_n = \frac{1-\alpha}{\alpha - \alpha^{K+1}} \sum_{k=1}^{K} \alpha^k y_{n-k} \tag{6.2}$$

Assuming that the nonlinearity acts instantaneously, further inspection of figure 6.1 indicates that

$$y_n = gf(x_n) \tag{6.3}$$

and combining (6.2) and (6.3) gives

$$x_{n+1} = \frac{1-\alpha}{1-\alpha^K} gf(x_n) + \frac{1-\alpha}{1-\alpha^K} g \sum_{k=1}^{K-1} \alpha^k f(x_{n-k}) \tag{6.4}$$

and for the special case where $K = \infty$

$$x_{n+1} = (1-\alpha)gf(x_n) + \alpha x_n \tag{6.5}$$

Note that the summation on the right handside of (6.4) vanishes when $K = 1$, in which case the loop reduces to the simple form recalled in the first sentence of this Chapter.

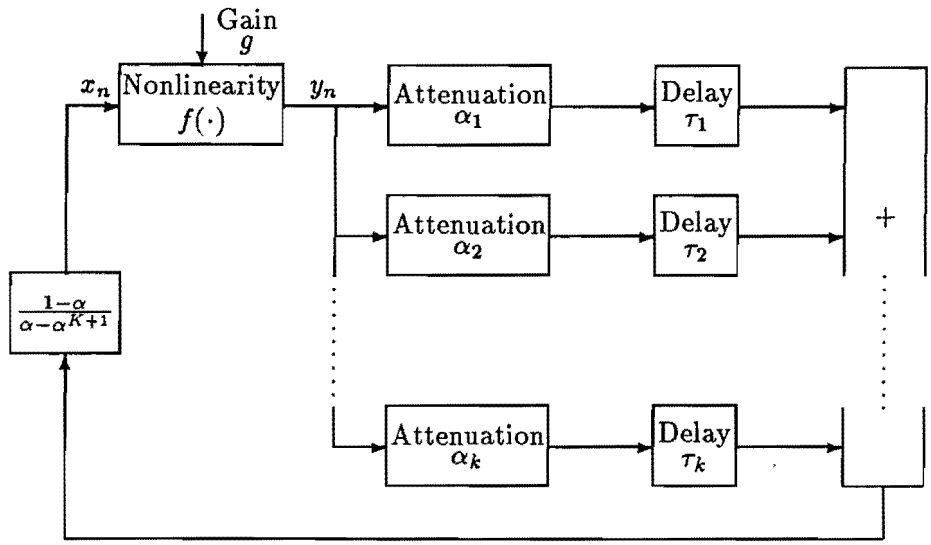


Figure 6.1: Recursive loop with memory. The input to the nonlinearity $f(\cdot)$ is formed from the sum of the output of the K parallel attenuating/delay paths. The individual attenuations are integer powers of a minimum attenuation α, and the individual delays are multiples of a minimum delay τ_0.

The six bifurcation diagrams (refer to §1.1.6) shown in figure 6.2 highlight the qualitative behaviour of (6.4). Each diagram in figure 6.2 represents a myriad of bifurcations, culminating ultimately in chaos. The six curves shown in figure 6.3 plot the largest Lyapunov exponent λ (refer to §1.1.12) against the gain constant g. The domain of g for each plot is the same as for the bifurcation diagrams shown in figure 6.2. In each case the largest Lyapunov exponent is greater than zero for a subset of the domain of g. The recursive loop with memory is chaotic (in the sense discussed in the last paragraph in §1.1.8) for those values of g where the Lyapunov exponent is greater than zero.

It is now appreciated that a great variety of noise-like processes can be explained purely deterministically (cf. Hao 1984; Cvitanovic 1984). Chua et al. (1986) has shown that a particular electronic oscillator (which now bears his name), without any explicit source of noise within it, can generate an apparently noisy signal (cf. §6.6). Also discrete signals generated by interconnected recursive loops can appear effectively stochastic. Such interconnected recursive loops can generate a wide range of spectral colourings and probability density functions (refer to Chapter 5). In addition, the fact that chaos persists in the discrete time oscillator described in the opening paragraphs of this Chapter suggests that it makes sense to try to discover to what extent noise within high quality sinusoidal oscillators can be explained in terms of deterministic chaos. This Chapter reports on results of investigations attempting to unravel this question.

Electronic oscillators are at the heart of accurate time-keeping and low-noise communication systems (Chi 1966; Jespersen et al. 1972; Rutman 1978; Robins 1984). Without highly stable sinusoidal oscillators it would, for example, be impracticable to modulate baseband information on to carriers for multi-channel communication

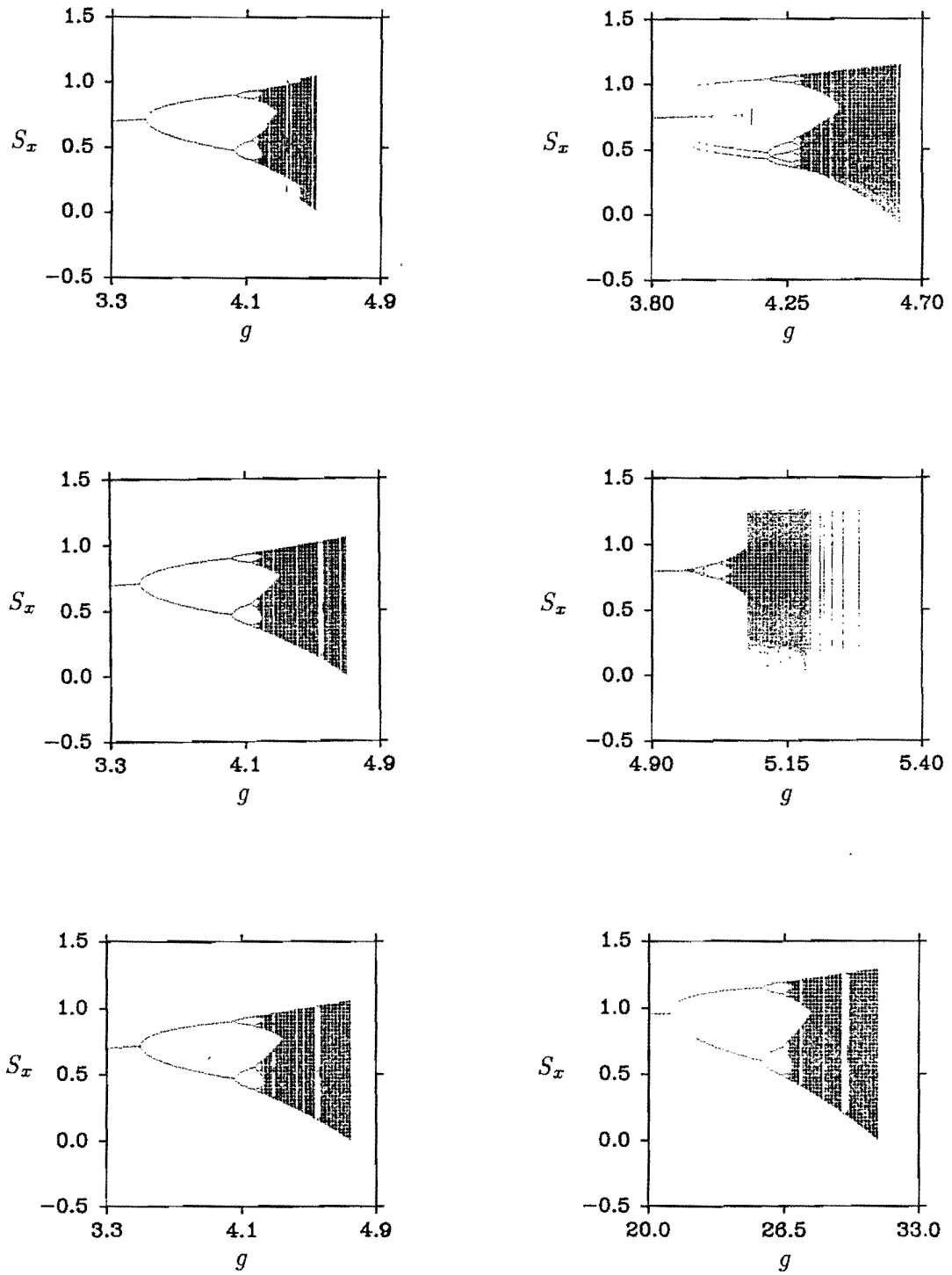


Figure 6.2: Bifurcation diagrams for the recursive loop with memory. Each diagram is constructed by plotting 1000 values of the sequence $S_x = \{x_n : n = 0, 1, \cdots\}$ against the gain constant g. The domain of g is shown on the x axis of each bifurcation diagram. The top, middle and bottom pair of diagrams are for $K = 2$, $K = 3$ and $K = \infty$ respectively. The left and right hand columns are for $\alpha = 0.2$ and $\alpha = 0.9$ respectively.

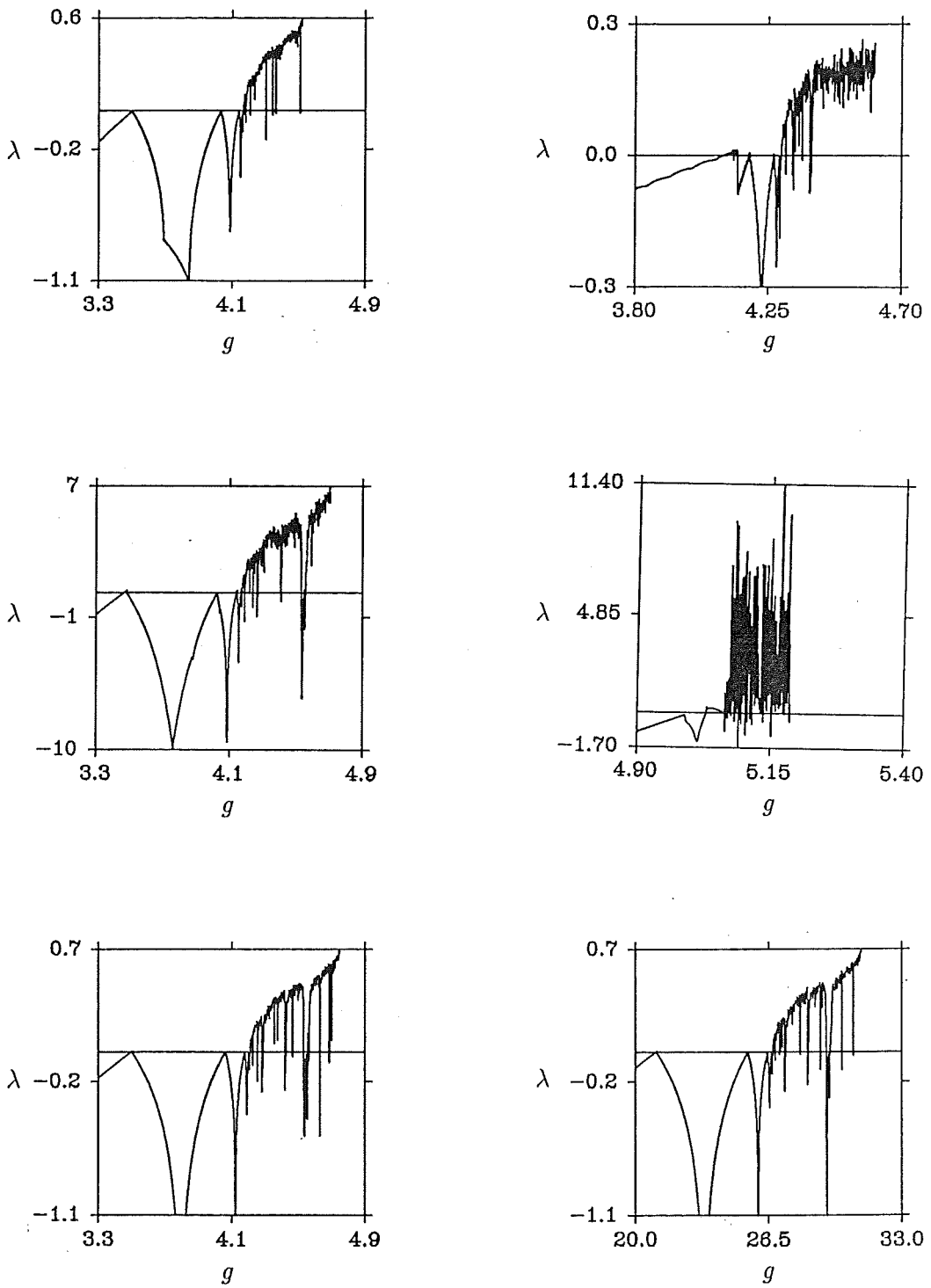


Figure 6.3: The largest Lyapunov exponent plotted against the gain constant g for the recursive loop with memory. The domain of g is shown on the x axis of each plot. The top, middle and bottom pair of plots are for $K = 2$, $K = 3$ and $K = \infty$ respectively. The left and right hand columns are for $\alpha = 0.2$ and $\alpha = 0.9$ respectively.

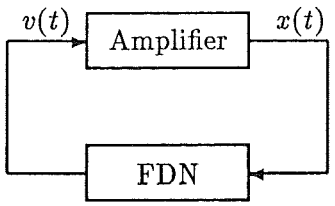


Figure 6.4: A basic oscillator loop, with $v(t)$ and $x(t)$ being respectively, the input signal to and output signal from the amplifier.

systems. Increasing demand on the frequency spectrum continually reduces the permissible tolerance on the spectral purity of such oscillators (Rutman 1978). If a carrier with unwanted amplitude or phase noise is modulated either in amplitude or phase by an information carrying baseband signal then the signal quality or error rate of the communication link is inevitably worsened. In practice, spurious phase noise is of more concern because it is usually of a much higher level than spurious amplitude noise (Robins 1984, page 47).

Any sinusoidal oscillator contains two essential parts; an *amplifier*, and a *frequency determining network* (FDN) (Beurle 1956; Hafner 1966; Robins 1984). Both the amplifier and the FDN have an input and an output. When the output of the amplifier is connected to the input of the FDN, and the output of the FDN is connected to the input of the amplifier, an *oscillatory loop* is formed. The amplifier forms the *active part* (i.e. it supplies the energy driving the signal around the loop) of the oscillator loop, and the FDN forms the *passive part* (i.e. removes power from the loop, although usually only in small amounts) of the oscillator loop. The amplifier is made up of at least one active circuit component (e.g. field effect transistor, tunnel diode). Figure 6.4 depicts a basic oscillator loop, with $v(t)$ and $x(t)$ being respectively, the input signal to and output signal from the amplifier. The output of the amplifier is assumed to appear instantaneously with the input $v(t)$, implying that all delays (such as transit times of charge carriers) are incorporated into the FDN. The output is expressed in terms of the input by

$$x(t) = gf(v(t)) \tag{6.6}$$

where $f(\cdot)$ is a normalized gain function in the sense that $\lim_{z\to 0} f(z) = z$. The constant g is thus the *low-level gain* (i.e. it is the gain for an infinitesimal input). The gain g together with the function $f(\cdot)$ is called the *amplifier characteristic*. It is useful to define the average gain by

$$< g >= \frac{1}{2T}\int_{-T}^{T} \frac{x(t)}{v(t)} dt \tag{6.7}$$

where T is appreciably longer than the reciprocal of the lowest significant frequency in the spectrum of $v(t)$ or $x(t)$.

The total loss around the oscillator loop is denoted O_ℓ, and the product $< g > O_\ell$ is termed the *loop gain*. The quantities $< g >$ and O_ℓ are in general complex and functions of frequency. If the real part of the loop gain is positive then the oscillatory loop is said to exhibit *positive feedback*. If there is at least one frequency for which

the loop gain is purely real and greater than or equal to unity, the oscillator loop is said to be *oscillatory* (i.e. it forms an oscillator). An oscillator generates a signal from no external inputs. This signal is termed the *oscillator signal.* The output of an oscillator is usually taken from the output of the FDN, and is termed the *output signal.* An oscillator is a *sinusoidal oscillator* if it generates a signal whose spectral content is clustered narrowly around a single dominant frequency. The set of frequencies for which $< g > O_\ell = 1$ is here denoted by f_0 (note that for a sinusoidal oscillator f_0 possesses a single member). It is convenient to consider the oscillator signal as the addition of two separate signals; the *carrier* and *noise.* The sinewave whose frequencies are f_0, $2f_0$, $3f_0$, etc., are termed respectively the carrier, second harmonic of the carrier (or just second harmonic), third harmonic of the carrier (or just third harmonic), etc. In practice the carrier and its harmonics have a bandwidth equal to the passband of the measuring instrument which is used to characterise them. The remainder of the oscillator signal after the carrier has been removed, is termed noise. The *spectral purity* of the oscillator signal is characterised by the carrier to noise ratio S_{f_0} (expressed in dB) which is given by

$$S_{f_0} = 10 \log \frac{C_0}{N} \tag{6.8}$$

where C_0 is the power of the carrier and N is the power of the noise. The carrier to n^{th} harmonic ratio S_{nf_0} (expressed in dB) is given by

$$S_{nf_0} = 10 \log \frac{C_0}{C_n} \tag{6.9}$$

where C_n is the power of the n^{th} harmonic of the carrier. Since only harmonics up to the third are considered in this Chapter, the notation can be simplified. The set of three ratios $\{S_{f_0}, S_{2f_0}, S_{3f_0}\}$ is denoted S and is termed the set of carrier ratios.

Conventional wisdom states that an oscillator begins oscillating by amplifying noise already present within the oscillator loop (Beurle 1956; Hafner 1966; Robins 1984). These oscillations build up until the oscillator signal reaches its full amplitude. This implies that after being initially greater than $1/O_\ell$, $< g >$ must reduce so as to equal $1/O_\ell$ (actually $< g >$ reduces to a value slightly less than $1/O_\ell$, since in reality noise supplies some energy to the oscillator signal) when the oscillator signal is at full amplitude. It is convenient to group sinusoidal oscillators into two classes, depending on the mechanism which causes $< g >$ to reduce as the oscillator signal amplitude increases. These two classes of oscillator are termed; *nonlinear gain*, and *linear gain-controlled.* For both classes, the average gain $< g >$ of the amplifier is a nonlinear function of the oscillator signal.

The departure of the amplifier characteristic from linearity is here quantified by

$$d = \int_{-\alpha}^{\alpha} \mid g f(z) O_\ell - 1 \mid dz \tag{6.10}$$

where α is the largest value of $\mid v(t) \mid$ when the oscillator signal has reached its full amplitude. The quantity d is a convenient measure of the *severity* of the amplifier nonlinearity. Because a signal passed through any nonlinearity undergoes *intermodulation* (i.e. modulation of the signal by itself), one sees that the concept of the severity of a nonlinearity is of considerable practical significance. The amount of intermodulation (and therefore the degree of signal degradation) increases with the

severity. For this reason, the *quality* of a sinusoidal oscillator is considered to be inversely related to the severity of the amplifier nonlinearity. A linear gain-controlled oscillator is generally of greater quality than a nonlinear gain oscillator.

By definition, $x(t)$ is a nonlinear function of $v(t)$ for a nonlinear gain oscillator (i.e. $x(t) = gf(v(t))$ where $f(\cdot)$ is a nonlinear function). Models of two nonlinear gain oscillators are discussed in detail in §6.2. The amplifier characteristic of one of these oscillators is the soft limiter $\tanh(gv(t))$, and typifies the most common type of nonlinear gain oscillator. The amplifier characteristic of the other oscillator is the cubic nonlinear function $gv(t)(1 - v^2(t))$, and models a tunnel diode. By definition, $x(t)$ is a linear function of $v(t)$ for a linear gain-controlled oscillator (i.e. $x(t) = gv(t)$, where g is a function of the amplitude of the envelope of $v(t)$ and $f(z) = z$). A particular model for a linear gain-controlled oscillator is also discussed in §6.2.

Two effects which tend to be neglected when sinusoidal oscillators are analysed are signal delay around the oscillator loop, and variations in signal delay and circuit parameters as a function of the oscillator signal. Models characterising these two effects are presented in §6.3. In practice, transit times of charge carriers across active circuit components constitute most of the signal delay around an oscillator loop (Moll 1955; Terman 1982). Furthermore, stray capacitances within amplifiers are the parameters whose values are usually most dependent on the level of the oscillator signal (Holt 1978). These effects increase the apparent noise content of the oscillator signal by introducing erratic intermodulation of the oscillator signal.

The consequences of introducing signal-dependent delay and capacitances into the oscillator models presented in §6.2 are analysed in §6.5 and §6.4. The discussion of nonlinear gain oscillators in §6.4 is split into three subsections. §6.4.1 examines the behaviour of a nonlinear gain oscillator having a soft limited amplifier characteristic. The consequences of introducing a signal dependent delay and capacitance is to raise the level of the harmonic content of the oscillator output signal. §6.4.2 examines the behaviour of a nonlinear gain oscillator whose amplifier is a tunnel diode. With an arbitrary signal delay, and a sufficiently large gain chaotic behaviour occurs (*cf.* Kitano *et al.* 1983). It is found that introducing signal-dependent delay and capacitance degrades the spectral purity by widening the spectral distribution of the oscillator output signal. §6.4.3 discusses two conditions which are sufficient for a nonlinear gain oscillator to exhibit deterministic chaos. It is conjectured in §6.4.3 that almost all nonlinear gain oscillators might satisfy these conditions, and as a consequence exhibit deterministic chaos. A linear gain-controlled oscillator is examined in §6.5. Similar conclusions are reached for this oscillator as for the soft limited nonlinear gain oscillator studied in §6.4.1.

The noise performance of a typical real-world sinusoidal oscillator seems to be somewhat worse than established approaches to noise analysis of theoretical models of such oscillators would suggest (Rutman 1978; Robins 1984, page 63). The conventional view to noise analysis, which has been formulated over the last 30 years or so, is outlined in §6.1. The purpose of §6.1 is to highlight the assumptions behind this approach to noise analysis, and to explain the noise mechanisms conventionally considered responsible for the non-spectral purity of sinusoidal oscillator signals.

The computer simulations presented in §6.6 show that Chua's circuit is sensitive to circuit details. Relatively minor circuit alterations can inhibit the onset of de-

terministic chaos. §6.6 demonstrates that deterministic noise-like effects persist in generalisations of Chua's circuit which incorporate signal delays and signal-dependent circuit component variations.

New results are presented in §6.4, §6.5 and §6.6. The conclusions that can be drawn from the material presented in this Chapter, together with suggestions for further work, are included in Chapter 9.

6.1 Review of Established Treatments of Oscillator Noise

The non-spectral purity of the output signal from a high quality oscillator is conventionally considered to arise from two sources; excess low frequency noise modulated on to the carrier, and thermal noise existing at frequencies adjacent to the carrier (*cf.* Jespersen 1972; Rutman 1978; Robins 1984). Due to the amplifier nonlinearity, signals present within the amplifier always suffer intermodulation. Intermodulation between the carrier and thermal noise close in frequency to the oscillator signal, and between thermal noise alone, produce unwanted additional noise. In high quality oscillators these intermodulation components are effectively negligible (Hafner 1966). However, intermodulation of the carrier and any excess low frequency noise which is present produces noise components of significant amplitude close in frequency to the carrier. When analysing the noise content of an oscillator output signal, the following assumptions are usually made (Beurle 1956; Hafner 1966; Robins 1984):

- The amplifier characteristic is linear (i.e. has an ideal linear amplifier), although Hafner (1966) does analysis an oscillator with a nonlinear amplifier characteristic.
- The exact circuit configuration is unimportant as far as the derivation of the carrier noise performance equations are concerned.
- The output signal is only due to noise amplified, via positive feedback, and filtered.
- Effects due to signal-dependent variations of circuit components and delay are negligible.

I have choosen to analyse the simple circuit shown in figure 6.5. The FDN is formed by the resistor R and a tank circuit consisting of the inductor L and the capacitor C. All circuit losses are included in R, and the thermal noise voltage v_n is generated within R. Reactive components at the amplifier input and output are assumed to have been tuned out by a circuit having a quality factor Q which is low compared to that of the FDN (*cf.* Hafner 1966; Skilling 1974).

Robins (1984) shows that the power spectrum of the phase noise present on the oscillator output signal $v(t)$ when thermal noise only is present within the oscillator loop is

$$\frac{N}{C_0} = \frac{FkT}{C_0} \frac{1}{8Q_e^2} \left(\frac{f_0}{\delta f}\right)^2 \tag{6.11}$$

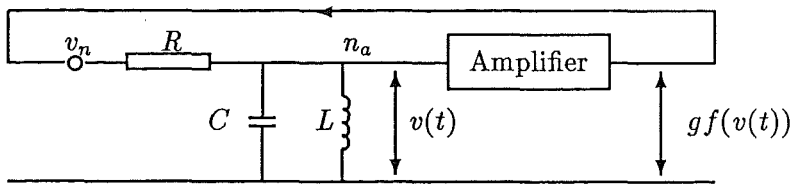


Figure 6.5: The oscillator circuit chosen for analysis. The FDN is formed by the resistor R and the tank circuit consisting of the inductor L and the capacitor C. The thermal noise voltage v_n is generated within R. Reactive components at the amplifier input and output are assumed to have been tuned out by a circuit having a Q low compared to that of the FDN.

where N is the phase noise power in the output signal, C_0 is the carrier power, F is the noise factor of the amplifier, k is Boltzmans constant, T is temperature in Kevin, δf is the frequency offset from f_0, and Q_e is the effective Q of the FDN tank circuit in the oscillator loop (effective Q is explained in the next section). Inspection of (6.11) reveals that the phase noise power is proportional to the inverse square of the offset frequency δf. However, (6.11) only holds for $\mid \delta f \mid$ greater than some transition frequency f_t, where the modulation components due to excess low frequency noise are insignificant. For $\mid \delta f \mid$ less than f_t the phase noise power is proportional to the inverse cube of δf. The value of the offset frequency at which the transition between $1/(\delta f)^2$ and $1/(\delta f)^3$ occurs can only be calculated if the excess low frequency noise performance and the amplifier characteristics are precisely known, which is not usually the case (Robins 1984). The established approach to noise analysis is unable to predict the level of the noise at frequencies very close to f_0. This approach also usually underestimates the noise level for frequencies greater than 0.1% away from f_0 (Robins 1984, page 63).

Established methods of oscillator noise analysis are unable to accurately predict the phase noise power in the oscillator output signal. The following sections consider to what extent ignoring contributions (if any) from deterministic chaotic mechanisms are responsible for this lack of precision.

6.2 Oscillator Models

This section develops the oscillator models that are analysed in depth in §6.4 and §6.5. These models are based on the circuit shown in figure 6.5. The FDN consists of a narrow band filter formed by the resistor R and the tank circuit consisting of the inductor L, and the capacitor C. The transfer function of the narrow band filter is given by (*cf.* Skilling 1974)

$$H(f) = \frac{jw/CR}{1/LC - w^2 + jw/CR} \tag{6.12}$$

The resonate frequency (specified in Hertz f_0, or radians per second w_0), of the tank circuit is given by

$$w_0 = \frac{1}{\sqrt{LC}} \tag{6.13}$$

and the 3dB bandwidth of the filter is $w_{bw} = w_2 - w_1$ where

$$\begin{aligned} w_1^2 &= w_0^2 + \frac{1}{2C^2R^2} - \frac{1}{CR}\sqrt{\frac{1}{4C^2R^2} + 4w_0^2} \\ w_2^2 &= w_0^2 + \frac{1}{2C^2R^2} - \frac{1}{CR}\sqrt{\frac{1}{4C^2R^2} + w_0^2} \end{aligned} \tag{6.14}$$

where w_1 and w_2 are the upper and lower 3dB cutoff radian frequencies respectively, and w_{bw} is the bandwidth of the narrow band filter in radians (*cf.* Skilling 1974). After applying Kirchhoff's current law to node n_a in figure 6.5, the following integro-differential equation is obtained

$$\frac{v(t) - gf(v(t))}{R} + C\frac{dv(t)}{dt} + \frac{1}{L}\int v(t)dt = 0 \tag{6.15}$$

After differentiating with respect to time and rearranging, the following differential equation results

$$\ddot{v}(t) + \frac{1}{CR}\left[1 - gf'(v(t))\right]\dot{v}(t) + \frac{1}{LC}v(t) = 0 \tag{6.16}$$

To simplify the notation, the analysis assumes that $f_0 = 1$, $C = 1$, $L = 1/w_0^2C$ and $R = 1$. With these simplifications (6.16) reduces to

$$\ddot{v}(t) + \left[1 - gf'(v(t))\right]\dot{v}(t) + w_0^2v(t) = 0 \tag{6.17}$$

which makes the bandwidth B of the narrow band filter 0.16Hz, and the quality factor Q (by definition $Q = f_0/B$) of the narrow band filter 6.28 (cf./ Skilling 1974). Because of positive feedback the effective bandwidth B_e and effective quality factor Q_e is (*cf.* Robins 1984, page 50)

$$B_e = B(1- < g > O_\ell) \text{ and } Q_e = \frac{Q}{1- < g > O_\ell} \tag{6.18}$$

where B_e is generally considerably less than B, and Q_e is generally considerably greater than Q (note that $< g > O_\ell$ is generally slightly less than unity; *cf.* Skilling 1974; Robins 1984, page 55). If the amplifier characteristic is the cubic nonlinear function $gv(t)(1 - v^2(t))$, where g specifies the low-level gain, then after a suitable variable transformation the oscillator studied by van der Pol (1934) is obtained:

$$\ddot{x} + \alpha(1 - x^2)\dot{x} + w_0^2x = 0 \tag{6.19}$$

where α is a constant. If the amplifier characteristic is the soft limited function $\tanh(gv(t))$, where g specifies the low-level gain, then a model for a soft amplitude limited oscillator is obtained

$$\ddot{v}(t) + \left[1 - g + g\tanh^2(\mathrm{gv(t)})\right]\dot{v}(t) + w_0^2v(t) = 0 \tag{6.20}$$

Equations (6.17) and (6.20) form the basis of the models for the nonlinear gain oscillators studied in depth in §6.4

Practical experience suggests that the noise performance of a typical nonlinear gain oscillator is adversely affected by the inescapable signal distortion introduced by the self-limiting action of the nonlinear gain amplifier. One way of countering

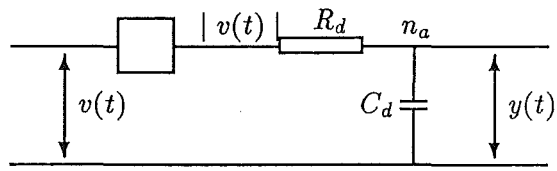


Figure 6.6: Circuit diagram of the 'model' envelope detector invoked for the linear gain-controlled oscillator. The output of the envelope detector controls the time-varying gain $g(t)$ of the amplifier. An estimate of the envelope of $v(t)$ is obtained by full wave rectifying $v(t)$ and passing the resultant through a single pole RC filter.

this is to continually adjust the gain so as to keep the average amplitude of the oscillator signal at some pre-chosen constant level. The amplitude of the envelope of the output signal $v(t)$ sets this gain. The amplifier (represented by the box labelled 'amplifier' in figure 6.5) in a linear gain-controlled oscillator consists of an envelope detector and a controlled gain $g(t)$. The envelope detector is conveniently modelled by a full wave rectifier followed by a single pole RC filter, as shown in figure 6.6. Applying Kirchhoff's current law to node n_a in figure 6.6 gives

$$\begin{gathered} \dot{y}(t) - \tfrac{1}{C_d R_d} y(t) + \tfrac{1}{C_d R_d} \mid v(t) \mid = 0 \\ g(t) = 1/y(t) \end{gathered} \tag{6.21}$$

Substituting $g(t)$ for g and setting $f(z) = z$ in (6.17) gives

$$\ddot{v}(t) + \dot{v}(t) - g(t)\dot{v}(t) + \dot{g}(t)v(t) + w_0^2 v(t) = 0 \tag{6.22}$$

The time constant of the single pole filter is set to 10× the length of one cycle of f_0, thus $R_d C_d = 10$. Equations (6.21) and (6.22) together form the basis of the model for the linear gain-controlled oscillator studied in depth in §6.5.

6.3 Signal-Dependent Transit Delay and Capacitance

Two phenomena important to the understanding of practical oscillators (especially at microwave frequencies) are ignored in the oscillator models developed in §6.2. These are the signal delay around the oscillator loop, and the variation of signal delay and circuit parameters as a function of the oscillator signal level (Hafner 1966; Terman 1982). Models characterising these effects are presented in this section. In practice, transit times of charge carriers across active devices constitute most of the signal delay around the oscillator loop (Moll 1955; Terman 1982). Furthermore, stray capacitance within amplifiers are the parameters whose values are most dependent on the level of the oscillator signal (Holt 1978).

The transit time for a charged carrier to cross an active device is a function of the electric field strength between its electrodes. The electric field in a vacuum tube is established between the cathode and anode (Terman 1982). Electrons travel across the free space between these electrodes. The electron acceleration is proportional to the electric field strength (Terman 1982, page 169). The transit time $t_r(t)$ for an

electron to travel the free space distance is given by

$$t_r(t) = \sqrt{\frac{ml^2}{qV}} \tag{6.23}$$

where m and q are the mass and charge of an electron respectively, and l and V are the distance and the potential difference between the cathode and anode respectively. In common cathode circuit configurations (the most common configuration; *cf.* Terman 1982) the anode potential is the addition of the oscillator signal voltage $v(t)$ and a DC bias voltage V_b. Thus the electron transit time becomes

$$t_r(t) = \sqrt{\frac{ml^2}{q[V_b + v(t)]}} \tag{6.24}$$

In semiconductor devices the charge carrier velocity (and not the acceleration) is proportional to the electric field strength between its electrodes (Moll 1955). The time taken by a charge carrier to transit a semiconductor is given by

$$t_r(t) = \frac{l^2}{\mu V} \tag{6.25}$$

where μ is the mobility of the charge carriers and l and V are the distance and the potential difference between the electrodes respectively. The output electrode consists of the oscillator signal voltage $v(t)$ superimposed on a DC bias voltage V_b. Thus the transit time of a charge carrier to travel across a semiconductor is given by

$$t_r(t) = \frac{l^2}{\mu[V_b + v(t)]} \tag{6.26}$$

The charged carrier transit time in both vacuum tubes and semiconductors consists of a signal-dependent delay superimposed on to a fixed delay. The signal delays specified by (6.24) and (6.26) are modelled here by a single lumped delay positioned immediately before the amplifier in the oscillator loop, as indicated in figure 6.7. Equation (6.26) can be expanded using the binomial theorem to give

$$t_r(t) = \frac{l^2}{\mu}\left[\frac{1}{V_b} - v(t) + \text{ higher order terms }\right], \tag{6.27}$$

and ignoring the higher order terms $t_r(t)$ can be characterised by

$$t_r(t) = t_d - \delta t_d v(t) \tag{6.28}$$

where t_d characterises the fixed length delay and δt_d characterises the proportion of $t_r(t)$ which is a function of the oscillator signal voltage $v(t)$.

In bipolar semiconductors the depletion and diffusion capacitances are functions of the electric field strength between the electrodes and the amount of current flow. Junction depletion capacitance C_J is given by (Holt 1978)

$$C_J = \frac{k}{(V_0 + V)^m} \tag{6.29}$$

where k is a constant depending on the area of the junction and impurity concentration levels, m is a constant depending on the distribution of impurities near the

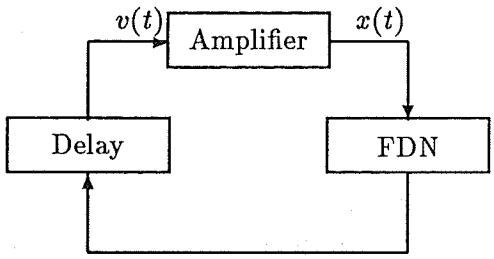


Figure 6.7: Block diagram of the oscillator loop with signal delay, which is modelled by a lumped component positioned immediately before the amplifier.

junction (m ranges from 1/3 to 4), V_0 is the depletion layer voltage with zero external voltage applied (0.7V for silicon, 0.2V for germanium) and V is the voltage between the electrodes (V is negative when the junction is forward biased). Since the output electrode of a semiconductor consists of the oscillator signal voltage $v(t)$ superimposed on a bias voltage V_b (*cf.* Terman 1982; Holt 1978), the resulting capacitance variation is given by

$$C_J = \frac{k}{(V_0 + V_b + v(t))^m} \tag{6.30}$$

The diffusion capacitance C_D is proportional to the forward biased current (Holt 1978). Diffusion capacitance is effectively zero when a junction is reversed biased, as is the case for the base-collector junction of a transistor operating in common emitter mode (the most common configuration; *cf.* Holt 1978). Thus depletion capacitance dominates.

The depletion capacitance (6.30) associated with a semiconductor device usually adds in parallel with the tank circuit capacitance C. Thus the total tank circuit capacitance C_T is given by

$$C_T = C + \frac{k}{(V_0 + v(t))^m} \tag{6.31}$$

Equation (6.31) can be simplified by assuming that $m = 1$ and applying the binomial theorem to give

$$C_T = C + k\left[\frac{1}{V_0} - v(t) + \text{ higher order terms }\right] \tag{6.32}$$

Again ignoring the higher order terms of the binomial expansion, C_T can be characterised by

$$C_T = C - \delta C v(t) \tag{6.33}$$

where C characterises a constant capacitance and δC characterises the proportion of C_T which is a function of the oscillator signal voltage $v(t)$.

6.4 Nonlinear Oscillator

This section contains three subsections. §6.4.1 and §6.4.2 examine the effect of introducing signal-dependent delay and capacitance into two conventional nonlinear

gain oscillators. Two conditions which are sufficient for a nonlinear gain oscillator to exhibit deterministic chaos are a signal delay around the oscillator loop and a suitable amplifier nonlinearity. These two conditions are studied in §6.4.3. It is conjectured that almost all nonlinear gain oscillators might satisfy such conditions, and as a consequence exhibit deterministic chaos. The results presented in this and the sections following are obtained by numerically solving the appropriate oscillator differential equation. Relevant computational considerations, and their impact on the associated software, are covered (in enough detail for the reader to appreciate their significance) in §6.7.

6.4.1 Soft Limited Oscillator

The nonlinear gain characteristics of most oscillators are of the soft limiting type. This subsection studies the effect of introducing signal-dependent delay and capacitance into a soft amplitude limited oscillator. Substituting equations (6.28) and (6.33) (which characterise signal-dependent delay and capacitance) into the soft limited oscillator characterised by (6.20) gives

$$\ddot{v}(t) + \frac{\dot{v}(t)}{RC_T} - \frac{1}{RC_T}\left[g - g\tanh^2(\mathrm{gv(t - t_r(t))})\right]\dot{v}(t - t_r(t)) + \frac{v(t)}{LC_T} = 0 \qquad (6.34)$$

which differs from the soft amplitude limited oscillators studied by most researchers in three important ways:

- It contains a fixed signal delay of duration t_d.
- It contains an extra signal delay which is a function of the oscillator signal $v(t)$.
- A portion of the circuit capacitance is a function of the oscillator signal $v(t)$.

The bifurcation diagrams (refer to §1.1.6) shown in figure 6.8(a)-(b) highlight the qualitative behaviour of numerical solutions of (6.34). Figure 6.8(a) is the diagram obtained by plotting against g the value of $v(t)$ whenever $\dot{v}(t) = 0$, for $t_d = 1$, $\delta t_d = 0$, $C = 1$, and $\delta C = 0$. Figure 6.8(a) shows that the soft amplitude limited oscillator undergoes a bifurcation at $g \approx 1$ (termed a Hoff bifurcation; *cf.* Mees and Chua 1979), and for $g >\approx 1$ the oscillator is oscillatory. Figure 6.8(a) shows that the amplitude of the oscillator signal $v(t)$ depends on the low level gain g. Figure 6.8(b) shows the diagram obtained by plotting the value of $v(t)$ whenever $\dot{v}(t) = 0$ against t_d, for $g = 3$, $\delta t_d = 0$, $C = 1$ and $\delta C = 0$. The effect of signal delay is to introduce a frequency dependent phase shift around the oscillator loop (*cf.* Holt 1978). The total phase shift around the loop at the frequency f_0 is always a multiple of 2π. Note that, while a delay of $1/f_0$ introduces a phase shift of 2π at the frequency f_0, it does not alter the frequency of the carrier. Figure 6.8(b) shows that for a range of delays in the neighbourhood of $t_d = 0.5$, $t_d = 1.5$ and $t_d = 2.5$ no oscillator signal is generated (this is indicated by the dotted horizontal line centred at $v(t) = 0$ in figure 6.8(b)). The reason for this is that these delays introduce phase shifts around the loop such that the real part of the loop gain for all frequencies is less than unity (i.e. the oscillator loop does not exhibit positive feedback). Figure 6.8(b) shows that the amplitude of the oscillator signal depends on the signal delay around the oscillator loop.

The seven graphs shown in figure 6.8(c)-(i) characterise the carrier to noise ratio S_{f_0} (solid curves), the carrier to second harmonic ratio S_{2f_0} (dashed curves) and the

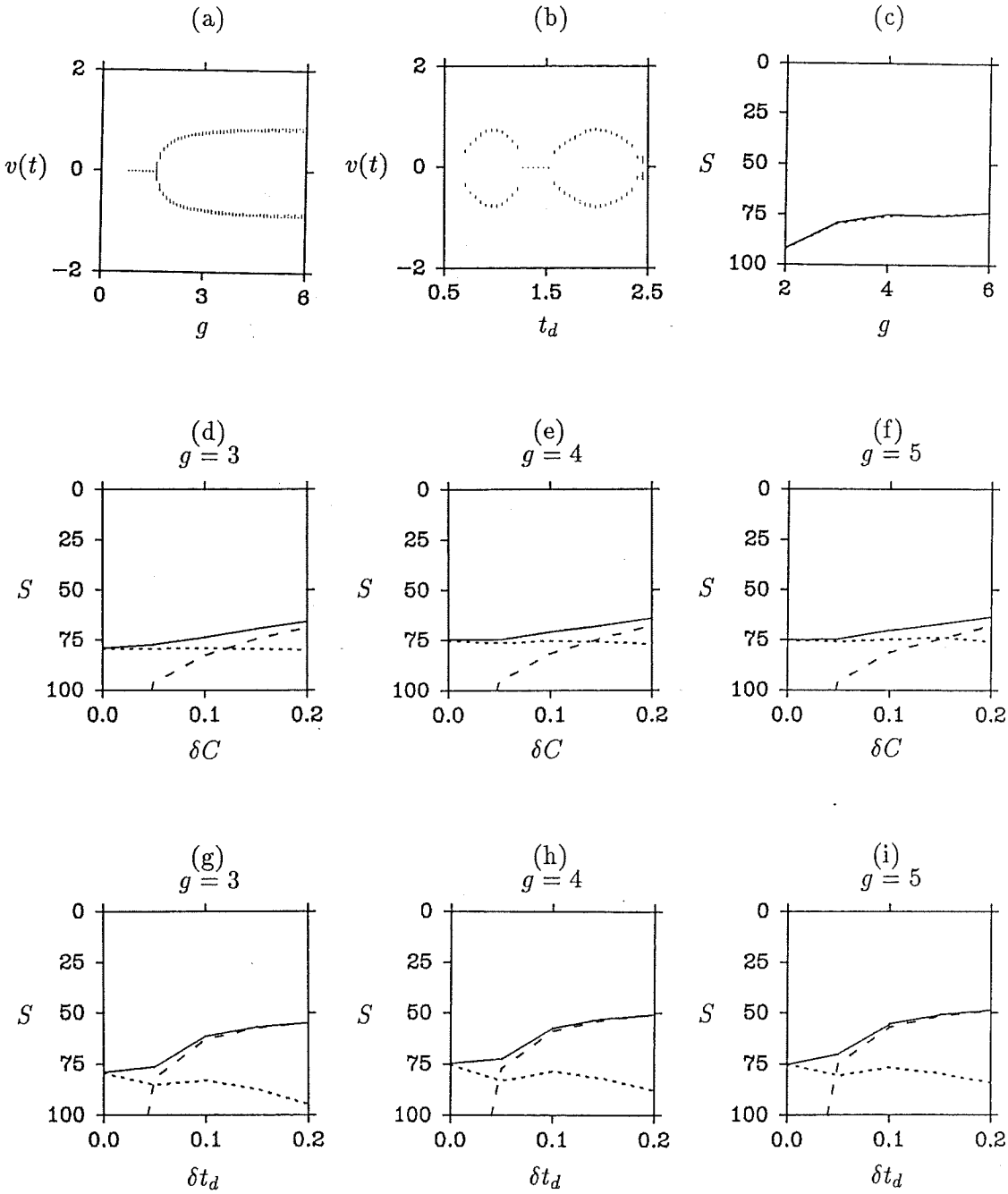


Figure 6.8: Results characterising the soft limited oscillator. **(a)** Bifurcation diagram obtained by plotting against g the value of $v(t)$ whenever $\dot{v}(t) = 0$, for $t_d = 1$, $\delta t_d = 0$, $C = 1$ and $\delta C = 0$. **(b)** Bifurcation diagram obtained by plotting against t_d the value of $v(t)$ whenever $\dot{v}(t) = 0$, for $g = 3$, $\delta t_d = 0$, $C = 1$ and $\delta C = 0$. The seven graphs labelled (c)-(i) characterise the carrier to noise ratio S_{f_0} (solid curves) the carrier to second harmonic ratio S_{2f_0} (dashed curves) and the carrier to third harmonic ratio S_{3f_0} (dotted curves) for the soft amplitude limited oscillator operating under the following conditions: **(c)** plot of S against g for $t_d = 1$, $\delta t_d = 0$, $C = 1$ and $\delta C = 0$, **(d)-(f)** plots of S against δC for $t_d = 1$, $C = 1$, $\delta t_d = 0$ and for the values of g specified on each plot, **(g)-(i)** plots of S against δt_d for $t_d = 1$, $C = 1$, $\delta C = 0$ and for the values of g specified on each plot.

carrier to third harmonic ratio S_{3f_0} (dotted curves), for the soft limited oscillator when operating under the conditions stated in the figure caption. Recall that the set of three ratios $\{S_{f_0}, S_{2f_0}, S_{3f_0}\}$ is denoted S and is termed the set of carrier ratios. Figure 6.8(c) plots S against g for $t_d = 1$, $\delta t_d = 0$, $C = 1$ and $\delta C = 0$. Since S_{f_0} and S_{3f_0} coincide (note that S_{2f_0} is greater than 100dB and is not plotted) in figure 6.8(c) then the non-spectral purity is entirely due to the presence of the third harmonic of the carrier. This is to be expected since the soft limited function (i.e. $\tanh(gv(t))$) is odd symmetric and distorts the negative and positive cycles of the oscillator signal waveform identically. This introduces odd harmonics (mostly third) of the carrier into the oscillator signal (Bracewell 1978).

Figure 6.8(d)-(f) plots S against δC for the values of g specified on each plot and for $t_d = 1$, $\delta t_d = 0$ and $C = 1$. The ratio S_{3f_0} is largely unaffected by a nonzero δC and remains approximately constant. However, the second harmonic rapidly increases with increasing δC and is the cause of the degradation in the spectral purity with increasing δC. Figure 6.8(g)-(i) plots S against δt_d for the values of g specified on each plot and for $t_d = 1$, $C = 1$ and $\delta C = 0$. The second harmonic increases (at a faster rate when compared with increasing δC) while the third harmonic actually decreases with increasing δt_d. Even quite small values of δt_d or δC can degrade the spectral purity by significant amounts. For low values of gain the effect of δt_d and δC on the spectral purity is less significant than for larger values of g.

Signal-dependent variations in circuit capacitance δC and signal delay δt_d causes the second harmonic of the carrier to increase significantly. This occurs because signal-dependent variations in capacitance and delay distorts the positive and negative cycles of the oscillator waveform unevenly (i.e. the waveform does not possess odd symmetry). Such distortion introduces even harmonics (in this case mostly second harmonics) of the carrier into the oscillator waveform. However, nonzero values of δC and δt_d) have much less effect on the level of the third harmonic. Signal-dependent variations in circuit capacitance and signal delay increase the level of intermodulation of the oscillator waveform (i.e. alter the shape of the oscillator waveform by introducing harmonics of the carrier) but do not introduce noise with a continuous frequency spectrum.

6.4.2 Tunnel Diode Oscillator

The circuit of a nonlinear gain oscillator whose amplifier is formed from a tunnel diode is shown in figure 6.9(b) (note that this circuit is equivalent to the circuit shown in figure 6.5). Typical I-V (current flow - terminal voltage) characteristics for a tunnel diode are shown in figure 6.9(a). If the tunnel diode is biased to the point indicated in figure 6.9(a) then the diode exhibits (for small signals) negative resistance. If a tunnel diode suitably biased to exhibit a negative resistance is connected across a tank circuit, and the negative resistance of the diode is of sufficient magnitude to cancel out the tank circuit loss (represented by R in figure 6.9(b)), then the circuit forms a sinusoidal oscillator. The tunnel diode I-V characteristics can be approximated by a cubic of the form $i(t) = gf(v(t)) = gv(t)(1 - v^2(t))$ where $i(t)$ is the current flow through the diode, $v(t)$ is the voltage across the diode and g is the low-level gain.

Applying Kirchhoff's current law at the node labelled n_a in figure 6.9(b) gives (6.16). Any charge carrier transit delay through the tunnel diode causes the current flowing through the tunnel diode to be delayed by $t_r(t)$ with respect to the applied

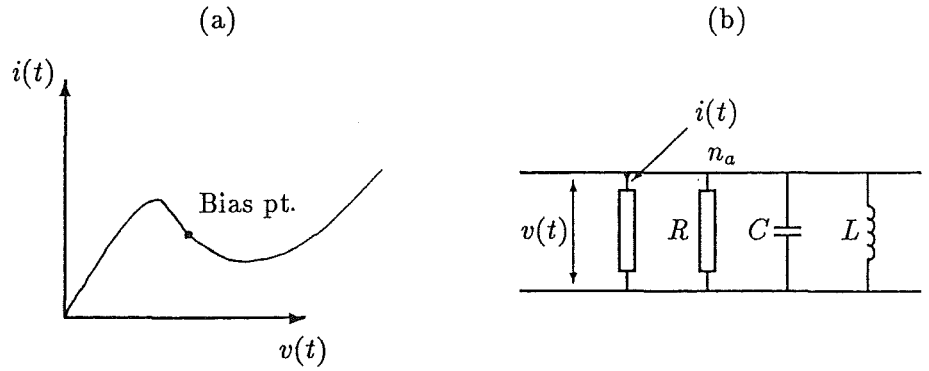


Figure 6.9: The tunnel diode oscillator. **(a)** Typical I-V characteristics for a tunnel diode. If the tunnel diode is biased to the point indicated then it exhibits (for small signals) negative resistance. **(b)** Equivalent circuit of the tunnel diode oscillator. The FDN is formed by the tank circuit, L and C. All circuit losses are included in R.

voltage (i.e. $i(t) = -gf(v(t - t_r(t)))$). Substituting the equations (6.28) and (6.33) (which characterise signal-dependent delay and capacitance) into (6.16) gives

$$\ddot{v}(t) + \frac{\dot{v}(t)}{RC_T} - \frac{1}{RC_T} f'[v(t - t_r(t))]\dot{v}(t - t_r(t)) + \frac{v(t)}{LC_T} = 0 \qquad (6.35)$$

which is similar to the van der Pol equation (6.19), but differs in three important ways:

- It contains a fixed signal delay of duration t_d.
- A portion of the signal delay is a function of the oscillator signal $v(t)$.
- A portion of the circuit capacitance is a function of the oscillator signal $v(t)$.

The bifurcation diagrams (refer to §1.1.6) shown in figure 6.10(a)-(b) highlight the qualitative behaviour of the numerical solution of (6.35). Figure 6.10(a) is the diagram obtained by plotting the value of $v(t)$ whenever $\dot{v}(t) = 0$ against g, for $t_d = 1$, $\delta t_d = 0$, $C = 1$, and $\delta C = 0$. Figure 6.10(a) shows that the tunnel diode oscillator undergoes at least two bifurcations, one at $g \approx 1$ (termed a Hoff bifurcation, *cf.* Mees and Chua 1979) and one at $g \approx 4$. For $g >\approx 1$ the oscillator is oscillatory, while for $g >\approx 4$, the waveform of $v(t)$ is complicated and is not a relatively pure sinusoid. Figure 6.10(a) shows that the amplitude and the complexity (i.e. the spectral purity) of $v(t)$ depends on the low level gain g. Figure 6.10(b) shows the diagram obtained by plotting the value of $v(t)$ whenever $\dot{v}(t) = 0$ against t_d, for $g = 3$, $\delta t_d = 0$, $C = 1$ and $\delta C = 0$. Figure 6.10(b) shows that for a neighbourhood of delays around $t_d = 0.5$, $t_d = 1.5$ and $t_d = 2.5$ no signal is generated (this is indicated by a dotted horizontal line centred at $v(t) = 0$ in figure 6.10(b)). This occurs for the same reasons as explained in §6.4.1. Figure 6.10(b) shows that the amplitude of the oscillator signal depends on the signal delay around the oscillator loop.

The six graphs shown in figure 6.10(c)-(h) characterise the carrier to noise ratio S_{f_0} (solid curves), the carrier to second harmonic ratio S_{2f_0} (dashed curves) and the carrier to third harmonic ratio S_{3f_0} (dotted curves) for the tunnel diode oscillator

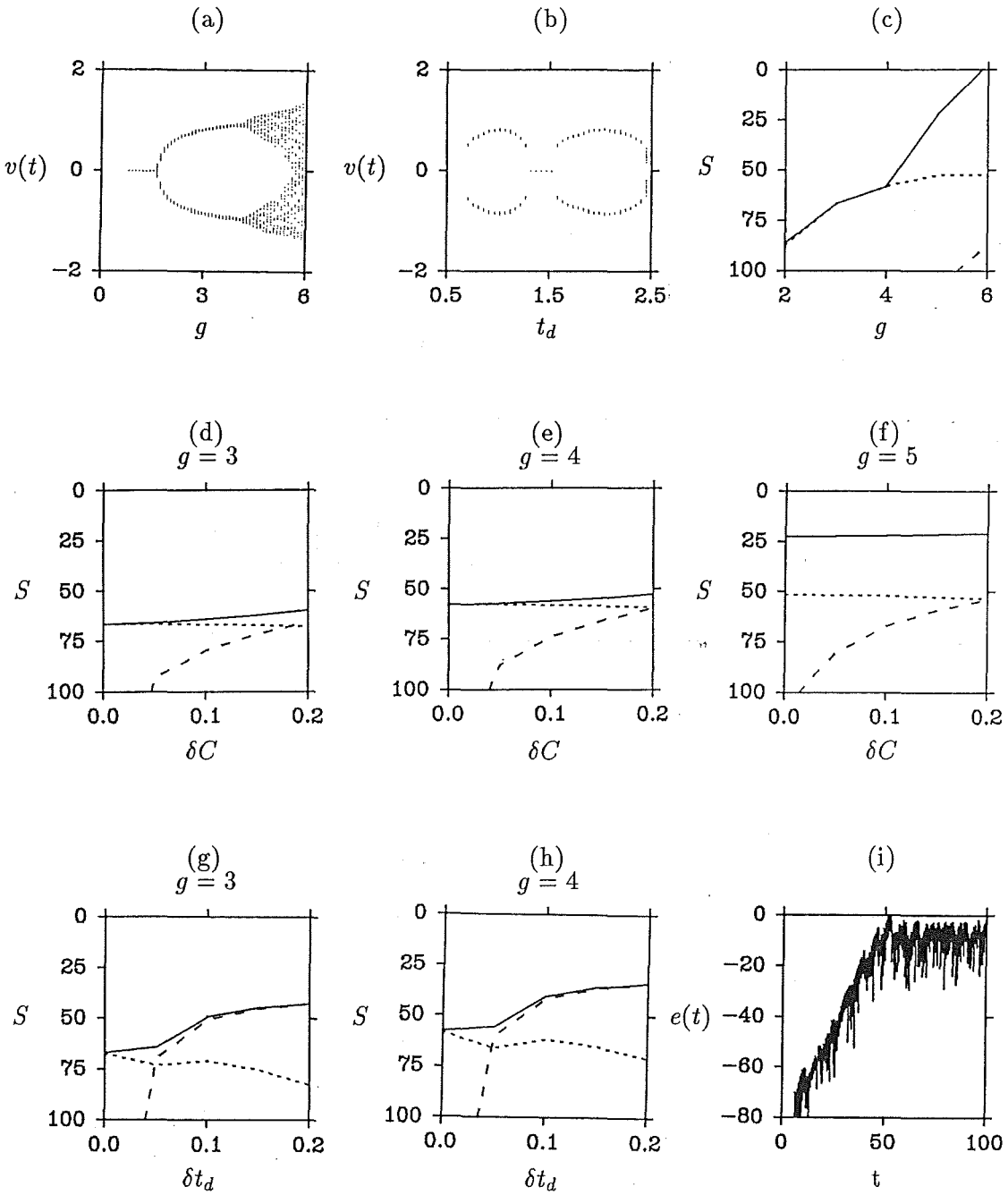


Figure 6.10: Results characterising the tunnel diode oscillator. **(a)** Bifurcation diagram obtained by plotting against g the value of $v(t)$ whenever $\dot{v}(t) = 0$, for $t_d = 1$, $\delta t_d = 0$, $C = 1$, $\delta C = 0$. **(b)** Bifurcation diagram obtained by plotting against t_d the value of $v(t)$ whenever $\dot{v}(t) = 0$, for $g = 3$, $\delta t_d = 0$, $C = 1$ and $\delta C = 0$. The six graphs labelled (c)-(h) characterise the carrier to noise ratio S_{f_0} (solid curves), the carrier to second harmonic ratio S_{2f_0} (dashed curves) and the carrier to third harmonic ratio S_{3f_0} (dotted curves) for the tunnel diode oscillator operating under the following conditions: **(c)** plot of S against g for $t_d = 1$, $\delta t_d = 0$, $C = 1$ and $\delta C = 0$, **(d)-(f)** plots of S against δC for $t_d = 1$, $\delta t_d = 0$, $C = 1$ and for the values of g specified on each plot, **(g)-(h)** plots of S against δt_d for $t_d = 1$, $C = 1$, $\delta C = 0$ and for the values of g specified on each plot. **(i)** Plot of the rate at which a perturbation of size 10^{-9} introduced at time 10 grows with time for $g = 6.0$, $t_d = 1$, $\delta t_d = 0$, $C = 1$ and $\delta C = 0$. The perturbation grows to the amplitude of $v(t)$ in 50 time units.

operating under the conditions stated in the figure caption. Figure 6.10(c) plots S against g for $t_d = 1$, $\delta t_d = 0$, $C = 1$ and $\delta C = 0$. Since S_{f_0} and S_{3f_0} coincide (note that S_{2f_0} is greater than 100dB and is not plotted) in figure 6.10(c) for $g <\approx 4$, the non-spectral purity is due entirely to the presence of the third harmonic of the carrier. This is to be expected since the nonlinear gain function of the tunnel diode is odd symmetric and introduces odd harmonics (mostly third) into the oscillator signal, as already explained in §6.4.1. For $g >\approx 4$, the spectral purity rapidly deteriorates with increasing g, while the level of the second and third harmonics increase only gradually. For $g >\approx 4$ the spectrum of the oscillator signal is continuous, having significant power at frequencies other than the carrier frequency or its harmonics (i.e. contains wideband noise). This wideband noise reduces the spectral purity of the oscillator signal significantly for $g > 4$.

Figure 6.10(d)-(f) plots S against δC for the values of g specified on each plot and for $t_d = 1$, $\delta t_d = 0$ and $C = 1$. The ratio S_{3f_0} is largely unaffected by δC and remains approximately constant. For figure 6.10(d)-(e), the second harmonic increases significantly with increasing δC and its presence is the cause of the degradation in the spectral purity. Figure 6.10(g)-(h) plots S against δt_d for the values of g specified on each plot and for $t_d = 1$, $C = 1$ and $\delta C = 0$. The second harmonic increases (at a faster rate when compared with increasing δC), while the third harmonic actually decreases, with increasing δt_d. As explained in the previous paragraph, for $g >\approx 4$ the oscillator signal contains wideband noise. The spectral content of this noise is largely unaffected by a nonzero δC or δt_d. This is demonstrated in figure 6.10(f) for $g = 5$ where the spectral purity is largely constant with increasing δC (note that the second harmonic increases with increasing δC in a similar way to that shown in the figures for $g < 4$). For $g < 4$, even quite small values of δt_d or δC can degrade the spectral purity to a significant extent. For low values of g the effect of a nonzero δt_d and δC on the spectral purity is less significant than for larger values of g. Signal-dependent variations in circuit capacitance δC and signal delay δt_d causes the second harmonic of the carrier to increase. This occurs for the same reasons as explained in §6.4.1.

Denote by $v(t)$ the trajectory obtained by numerically solving (6.35) for some particular initial condition. Denote by $\hat{v}(t)$ a second trajectory obtained by numerically solving (6.35) for the same initial condition, but with a perturbation of 10^{-9} added at time 10. Ten times the logarithm of the absolute different between $v(t)$ and $\hat{v}(t)$ normalized with respect to the largest value of $v(t)$ is here termed the error $e(t)$, i.e.

$$e(t) = 10 \log \left(\frac{| v(t) - \hat{v}(t) |}{\alpha} \right) \tag{6.36}$$

where α is the largest value of $v(t)$. Figure 6.10(i) plots $e(t)$ against t for $g = 6$, $t_d = 1$, $\delta t_d = 0$, $C = 1$ and $\delta C = 0$. This plot shows that the error increases with time until it is of similar magnitude to the amplitude of $v(t)$. Such behaviour is highly suggestive of deterministic chaos. In addition, the bifurcation diagram figure 6.10(a) is reminiscent of the bifurcation diagram of the logistic map shown in figure 1.8. This suggests that (6.35) is chaotic when $g >\approx 4.0$, $t_d = 1$, $\delta t_d = 0$, $C = 1$, $\delta C = 0$.

6.4.3 Conditions Sufficient for Deterministic Chaos

If the FDN in figure 6.7 is removed, the resulting oscillator loop is fully characterised by the one-dimensional recursive loop

$$x_{n+1} = gf(x_n) \tag{6.37}$$

where

$$x_{n+1} = v\left[(n+1)t_r(t)\right] \text{ and } x_n = v\left(nt_r(t)\right) \text{ for } n = 0, 1 \cdots \infty, \tag{6.38}$$

where $t_r(t)$ is the duration of the signal delay, and x_n is the value of $v(t)$ at the n^{th} time instance (i.e. at $t = nt_r(t)$). A necessary condition for a nonlinear oscillator to exhibit deterministic chaos is for the recursive loop (6.37) to exhibit deterministic chaos. Computational experience suggests that the FDN modifies but does not eliminate chaotic behaviour generated within a recursive loop. The nonlinear function $f(\cdot)$ characterising the amplifier of a nonlinear oscillator is usually assumed to be a smooth monotonically increasing function (e.g. a soft limiter). However, in reality this assumption is probably false. It is here proposed that on a minute scale $f(\cdot)$ is bumpy. This bumpyness may arise from the internal structure of the active device(s) that form the nonlinear amplifier. Mechanisms that could cause such a bumpyness in the case of the field effect transistor (FET) are proposed in the following paragraphs. A similar argument can be developed for the bipolar transistor.

The structure of a depletion mode metal oxide semiconductor (MOS) FET is shown in figure 6.11 and can be viewed as a voltage controlled resistance. The gate voltage controls the source-drain resistance (Holt 1978, page 219). A thin oxide insulator separates the semiconductor channel and the metal layer forming the gate. If the voltage between the gate and the source V_{gs} is made negative, channel electrons are repelled from under the gate. This forms a region depleted of electrons and is termed the depletion region. The depletion region forces the source-drain electron flow to become confined to a channel narrower than the size of the physical channel. This increases the resistance between the source and drain. The depletion region is not of uniform width along the channel. The width of the depletion region at a point along the channel depends on the voltage difference between that point and the gate. This voltage difference progressively increases along the length of the channel and results in the depletion region being wider towards the drain end.

It is unlikely that any semiconductor channel can be completely homogeneous. The electrical resistance of any small unit volume of the channel differs (however slightly) from other unit volumes in the same channel. A varying depletion region (induced by a varying gate voltage) covers and uncovers channel volumes of differing resistance. This causes the source drain current to change accordingly. Therefore a smoothly increasing gate voltage might not necessarily cause a smoothly increasing source-drain current. A plot of gate voltage versus source-drain current must necessarily be bumpy (i.e. not a smooth monotonically increasing function). This effect is here modelled by adding small bumps to a soft limiter function $f(\cdot)$. In particular, the behaviour of the one-dimensional recursive equation

$$x_{n+1} = \tanh(gx_n) + b\pi \sin\left(\frac{x_n}{b}\right) \tag{6.39}$$

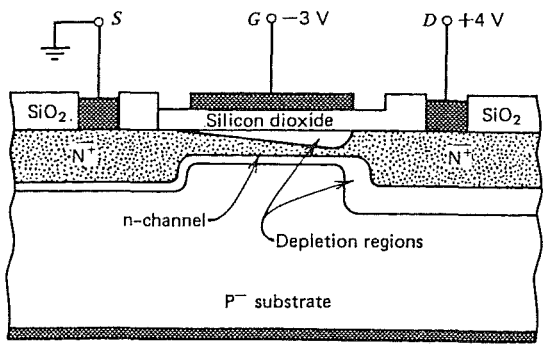


Figure 6.11: Structure of a depletion mode metal oxide semiconductor (MOS) FET. A thin oxide insulator separates the semiconductor channel from the metal layer forming the gate. If the voltage between the gate and the source is negative, electrons are repelled from under the gate. The depletion layer confines the source-drain electron flow, increasing the source-drain resistance.

is studied. The first term on the right hand side of (6.39) is a monotonically increasing soft limited function (the same function as used to model the amplifier characteristics of the soft limited nonlinear oscillator studied in §6.4.1) where g specifies the low-level gain. The second term on the right hand side of (6.39) models bumps, where b specifies their size. Recalling the final paragraph of §1.1.6, any non-chaotic recursive equation can be perturbed by an arbitrarily small scaled-down version of a chaotic recursive equation positioned at the equilibrium point(s) of a non-chaotic recursive equation to produce a chaotic recursive equation. Since the second term on the right hand side of (6.39) forms a chaotic map, then (6.39) is chaotic (provided the bumps are correctly positioned at the equilibrium point(s) of the first term on the right hand side of (6.39)). A small bump positioned at the equilibrium point of the first term on the right hand side of (6.39) generates chaos confined to a small region about the equilibrium of the first term. Arbitrary perturbations do not cause major observable changes in the behaviour of a non-chaotic recursive equation. However, bumps do provide a mechanism by which deterministic noise may arise within sinusoidal oscillators.

The left hand diagram in figure 6.12 shows the bifurcation diagram for (6.39) obtained by plotting 1000 values of the sequence $S_x = \{x_n : n = 0, 1, \cdots\}$ against each value of b for $g = 3$. The bifurcation diagram consists of a number of thick bands connected by rough sloping lines, which indicate chaotic and regular behaviour respectively. As b is increased, not only the size but also the distance between bumps increases (i.e. the position of the bumps on the soft limiter characteristic expand). Since the positions of the bumps determine whether (6.39) is chaotic or not, the bifurcation diagram consists of regions of chaotic behaviour separated by regions of regular behaviour. It is the bump size that is of interest here, and so only those values of b for which (6.39) is chaotic are considered further. The right hand curves in figure 6.12 plot S against values of b for which (6.39) is chaotic. The second harmonic is 100dB down on the carrier, and the third harmonic is constant at approximately 80dB down on the carrier, for the range of b values considered. However, the carrier

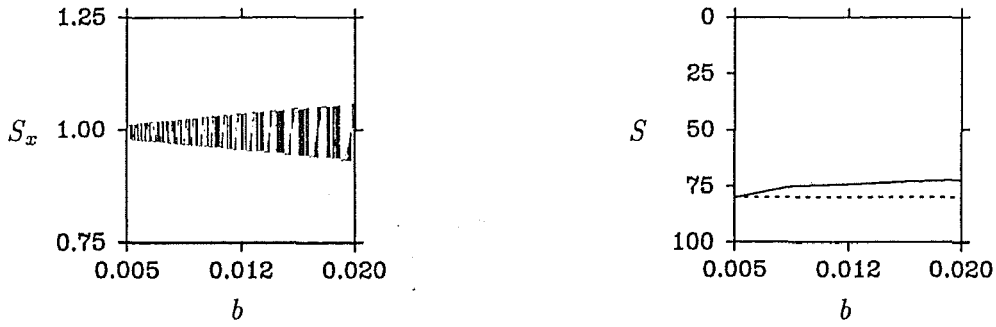


Figure 6.12: Results characterising an oscillator with a bumpy gain characteristic. The left hand diagram is a bifurcation diagram of (6.39) obtained by plotting 1000 values of the sequence $S_x = \{x_n : n = 0, 1, \cdots\}$ against each value of b for $g = 3$. The right hand curve is a plot of S against b for values of b for which (6.39) is chaotic.

to noise ratio S_{f_0} decreases with increasing b (i.e. the size of the bumps) indicating that the amplitude of the chaos is increasing. Even small bumps can reduce the spectral purity by a significant amount.

The results presented in figure 6.12 demonstrate that the existence of small bumps (on an amplifier characteristic) can cause deterministic chaos to arise within a nonlinear gain oscillator. However, the existence of such bumps on real-world oscillator gain characteristics has yet to be confirmed. The literature reveals nothing about such bumps. This may indicate that bumps, if they exist at all, are so small that they are difficult (or nearly impossible) to detect experimentally and/or the existence of such bumps have been hitherto considered unimportant for inclusion in oscillator models. The essential point is that, while the bump size determines the amplitude of the ensuring deterministic chaos, a suitable bump of any size results in deterministic chaos.

6.5 Linear Gain-Controlled Oscillator

By definition, $x(t)$ is a linear function of $v(t)$ for a linear gain-controlled oscillator, i.e. $x(t) = g(t)v(t)$, where $g(t)$ is the time-varying gain (refer to figure 6.6). Equations (6.21) and (6.22) characterise the linear gain-controlled oscillator. This section studies the effect of introducing the signal-dependent delays and capacitance variations discussed in §6.3 into a particular linear gain-controlled oscillator. Incorporating (6.28) and (6.33) into (6.21) and (6.22) gives

$$\begin{aligned} \dot{y}(t) + \tfrac{y(t)}{10} - \tfrac{|v(t)|}{10} &= 0 \\ g(t) &= A/y(t) \\ \ddot{v}(t) + \tfrac{\dot{v}(t)}{RC_T} - \tfrac{1}{RC_T} g(t - t_r(t))\dot{v}(t) + \tfrac{1}{RC_T}\dot{v}(t - t_r(t))(t)v(t) + \tfrac{v(t)}{LC_T} &= 0 \end{aligned} \tag{6.40}$$

where A sets the amplitude of $v(t)$. Equation (6.40) differs from the formulations for the linear gain-controlled oscillators studied by other researchers in three important ways:

- It contains a fixed signal delay of duration t_d.

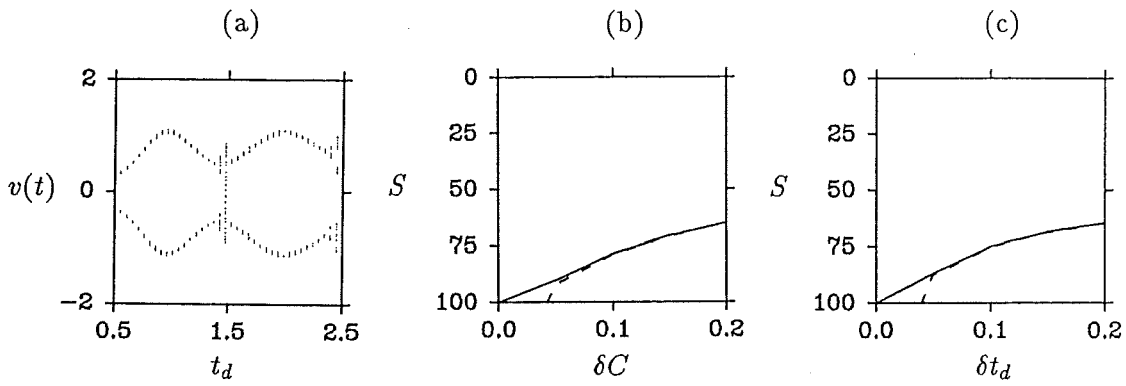


Figure 6.13: Results characterising the linear gain-controlled oscillator. (a) Bifurcation diagram obtained by plotting against t_d the value of $v(t)$ whenever $\dot{v}(t) = 0$, for $g = 3$, $\delta t_d = 0$, $C = 1$ and $\delta C = 0$. (b) Plot of S against δt_d for $t_d = 1$, $C = 1$ and $\delta C = 0$. (c) Plot of S against δC for $t_d = 1$, $C = 1$ and $\delta t_d = 0$.

- There is an additional signal delay which is a function of the oscillator signal $v(t)$.

- Part of the circuit capacitance is a function of the oscillator signal $v(t)$.

The bifurcation diagram shown in figure 6.13(a) highlights the qualitative behaviour of the numerical solution of (6.40). Figure 6.13(a) is obtained by plotting the value of $v(t)$, whenever $\dot{v}(t) = 0$, against t_d, for $g = 3$, $\delta t_d = 0$, $C = 1$ and $\delta C = 0$. Since the linear gain-controlled oscillator incorporates automatic gain control (agc), if the loss around the loop should change due to variations in circuit capacitance or signal delay, the agc can counter this by changing the amplifier gain $g(t)$. Figure 6.13(a) shows that the oscillator signal is of greatest amplitude when the signal delay has a duration equal to a multiple of a single cycle of the carrier, and is of least amplitude when the signal delay has a duration equal to an odd number of half cycles of the carrier. Although the amplitude of the oscillator signal is least when the signal delay has a duration of an odd number of half cycles, it is not zero as is the case for the nonlinear gain oscillators.

The carrier ratios S are characterised by the two graphs shown in figure 6.13(b)-(c). Figure 6.13(b) plots S against δt_d for $t_d = 1$, $C = 1$, $\delta C = 0$. Figure 6.13(c) plots S against δC for $t_d = 1$, $C = 1$, $\delta t_d = 0$. The spectral purity does not degrade to the same extent with increasing δt_d or δC as for the nonlinear gain oscillators. However, even quite small variations in t_d or C can degrade the spectral purity significantly. For both figure 6.10(b) and figure 6.10(c) the level of the third harmonic is 100dB down on the carrier, which is considerably less than the level present in the oscillator signal of the nonlinear gain oscillators studied in §6.4. The level of the third harmonic is less because the amplifier is linear and introduces very little waveform distortion. However, any signal-dependent variation in circuit capacitance or signal delay raises the level of the second harmonic significantly. Overall, the spectral purity of the soft limited oscillator is considerably better than that of the nonlinear gain oscillators.

6.6 Chua's Circuit

The first report of complicated dynamical behaviour from an oscillatory circuit first suggested by Chua (which now bears his name) was by Matsumoto (1984). Chua's circuit (also known as Chua's oscillator) has generated a lot of interest for the following reasons:

- It is the simplest autonomous electrical circuit which can become chaotic.
- It is the only physical system that has been demonstrated to be chaotic through all three of the accepted means of establishing engineering-scientific results, i.e. by computer simulation (Matsumoto 1984), laboratory experiments (Zhong and Ayrom 1985), and mathematical analysis (Chua *et al.* 1986).
- It exhibits an immense variety of dynamical phenomena, including many typical 'bifurcations' and 'scenarios preceding chaos' (Matsumoto *et al.* 1986b).

Chua's circuit, which is depicted in figure 6.14 is third order, reciprocal, and has one nonlinear element, a 5-segment piecewise-linear resistor. By applying Kirchhoff's voltage and current laws the circuit dynamics can be characterised by the ODE's,

$$\begin{aligned} C_1\frac{dv_{C_1}(t)}{dt} &= G(v_{C_2}(t)-v_{C_1}(t))-f(v_{C_1}(t)) \\ C_2\frac{dv_{C_2}(t)}{dt} &= G(v_{C_1}(t)-v_{C_2}(t))-i_L(t) \\ L\frac{di_L(t)}{dt} &= v_{C_2}(t) \end{aligned} \tag{6.41}$$

where $v_{C_1}(t)$, $v_{C_2}(t)$ and $i_L(t)$ denote the voltage across C_1, the voltage across C_2 and the current through L, respectively, and $f(v_{C_1}(t))$ is the 5-segment piecewise-linear function,

$$f(v_{C_1}(t)) = \begin{cases} m_2(v_{C_1}(t)-b_2)+b_2, & \text{if } v_{C_1}(t) > b_2 \\ m_1(v_{C_1}(t)-b_1)+b_1, & \text{if } b_1 < v_{C_1}(t) \leq b_2 \\ m_0 v_{C_1}(t), & \text{if } \mid v_{C_1}(t) \mid \leq b_1 \\ m_1(v_{C_1}(t)+b_1)-b_1, & \text{if } -b_2 \leq v_{C_1}(t) < -b_1 \\ m_2(v_{C_1}(t)+b_2)-b_2, & \text{if } v_{C_1}(t) < -b_2 \end{cases} \tag{6.42}$$

where m_0, m_1, m_2, b_1 and b_2 are constants. Figure 6.16 shows diagrammatically the 5-segment piecewise-linear function characterised by (6.42). Figure 6.15(a) shows 50 seconds of the waveform of $v_{C_1}(t)$; the waveform of $v_{C_2}(t)$ and $i_L(t)$ are similar in character (*cf.* Matsumoto 1987, figure 4). Observe that the waveform $v_{C_1}(t)$ consists of oscillations centred at two distinct voltage levels (i.e. ± 2 volts). The oscillations increase in amplitude until an abrupt disturbance occurs which either flips the oscillations onto the other level or returns them to the same level, whereupon the process repeats. The number of cycles constituting the oscillations between each abrupt disturbance is seemingly random. Chua *et al.* (1986) show that the dynamics of (6.41) reside on a strange attractor, which has been termed the double scroll. Figure 6.15(b)-(d) shows projections of the strange attractor onto the ($i_L(t)$, $v_{C_1}(t)$)-plane, ($v_{C_1}(t)$, $v_{C_2}(t)$)-plane and ($i_L(t)$, $v_{C_2}(t)$)-plane respectively for the

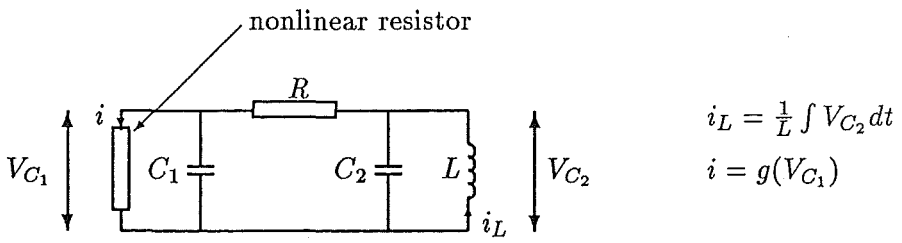


Figure 6.14: Circuit diagram of Chua's circuit, which is third order, reciprocal, and has one nonlinear element (a 5-segment piecewise-linear resistor) (Chua *et al.* 1986).

circuit parameter values

$$\begin{aligned} 1/C_1 = 9, 1/C_2 = 0.5, 1/L = 7, G = 0.7, \\ m_0 = -0.5, m_1 = -0.8, m_2 = 10.0, b_1 = 1, b_2 = 10 \end{aligned} \tag{6.43}$$

The plots shown in figure 6.15(a)-(d) are generated by numerically solving (6.41) for the parameters specified in (6.43) (for details of the numerical method used refer to §6.7).

Equation (6.41), with the parameter values specified by (6.43), has three unstable fixed points (refer to §1.1.1): one at the origin, one at the centre of the 'upper hole' of the strange attractor, and one at the centre of the 'lower hole' of the strange attractor (refer to figure 6.15(b)-(d)). A typical trajectory on the attractor rotates around one of the two outer fixed points, say the upper one, in a counterclockwise direction. After each rotation the trajectory moves further from the fixed point until a certain instant, after which there are two possibilities: 1) the trajectory moves back to a position closer to the fixed point and repeats a similar process, 2) the trajectory descends downward (with respect to the $v_{C_1}(t)$ - axis) in a spiral path to a position close to the lower fixed point. In the latter case, the behaviour is similar to that in the upper part of the attractor. The number of rotations a trajectory makes around a fixed point before it starts descending or ascending is seemingly random. The number of rotations it makes while it descends or ascends is also seemingly random.

This section assesses how robust the chaotic behaviour generated by Chua's oscillator is to circuit component values and to signal-dependent variations in component values. Matsumoto *et al.* (1985, page 801) reports that the strange attractor persists for at least the following parameter ranges:

$$\begin{aligned} 8.82 \leq 1/C_1 \leq 10.6 \text{ other parameters fixed at the values in (6.43)} \\ 0.43 \leq 1/C_2 \leq 1.08 \text{ other parameters fixed at the values in (6.43)} \\ 5.70 \leq 1/L \leq 7.13 \text{ other parameters fixed at the values in (6.43)} \\ 0.68 \leq G \leq 0.76 \text{ other parameters fixed at the values in (6.43)} \end{aligned} \tag{6.44}$$

However, Matsumoto *et al.* (1985) do not report on how sensitive the attractor is to signal delay within the circuit, the parameters characterising the nonlinearity, and signal-dependent variations in circuit capacitance. This section assesses these through the five bifurcation diagrams presented in figure 6.15(e)-(i). In each diagram all circuit parameters are set to the particular values specified in (6.43) except for

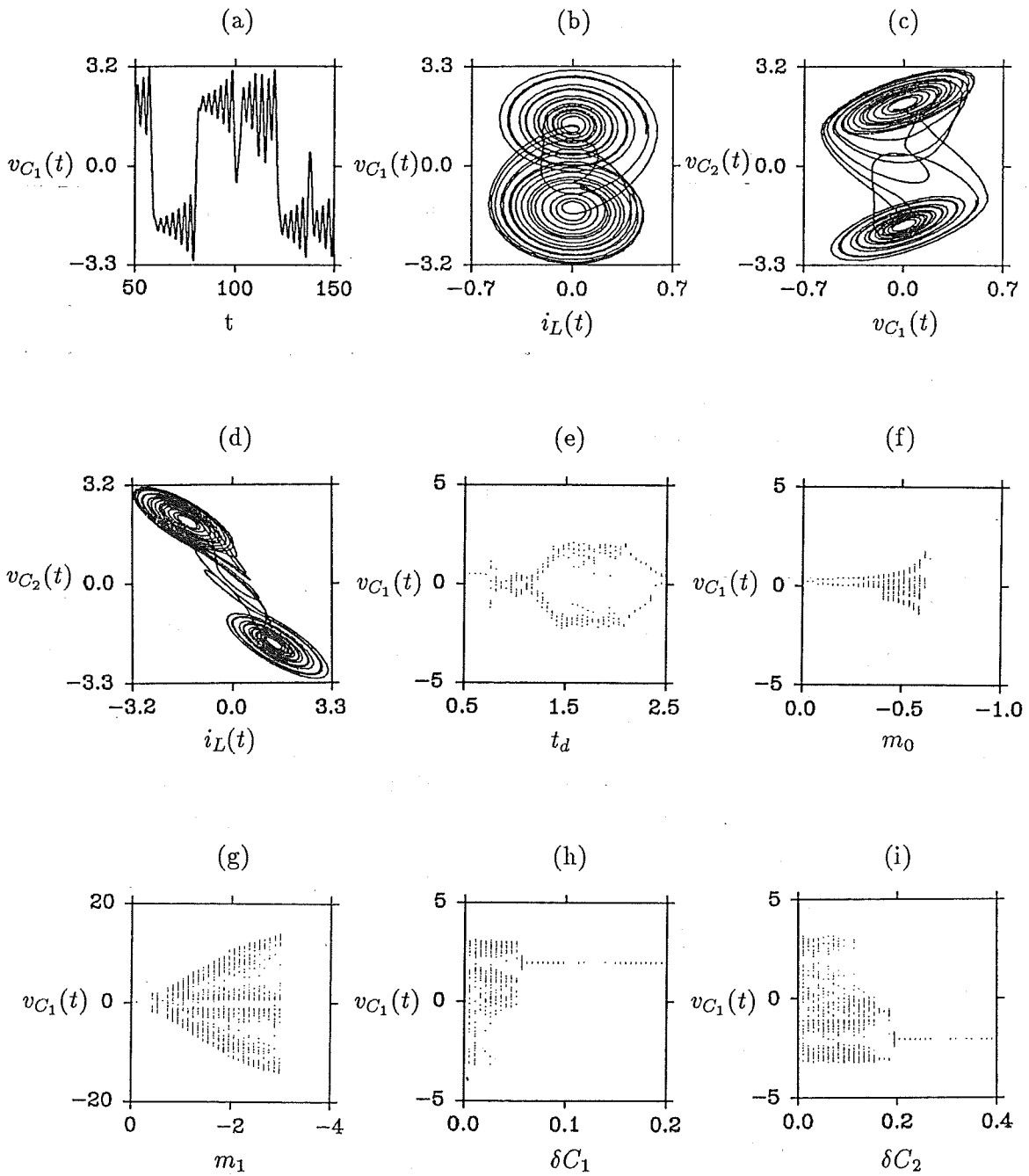


Figure 6.15: Results characterising Chua's circuit. **(a)** A plot of 50 seconds of $v_{C_1}(t)$. Projections of Chua's strange attractor onto **(b)** the (i_L, v_{C_1})-plane, **(c)** the (v_{C_1}, v_{C_2})-plane and **(d)** the (i_L, v_{C_2})-plane. Bifurcation diagrams obtained by plotting the values of $v_{C_1}(t)$ whenever $\dot{v}_{C_1}(t) = 0$ against: **(e)** t_d, for $m_0 = -0.5$, $m_1 = -0.8$, $\delta C_1 = 0$ and $\delta C_2 = 0$, **(f)** m_0, for $t_d = 0.0$, $m_1 = -0.5$, $\delta C_1 = 0$ and $\delta C_2 = 0$, **(g)** m_1, for $t_d = 0.0$, $m_0 = -0.8$, $\delta C_1 = 0$ and $\delta C_2 = 0$, **(h)** δC_1, for $t_d = 1.0$, $m_0 = -0.8$, $m_1 = -0.5$ and $\delta C_2 = 0$, **(i)** δC_2, for $t_d = 0.0$, $m_0 = -0.8$, $m_1 = -0.5$ and $\delta C_1 = 0$.

the one under study, whose range of values is shown on the x-axis of each bifurcation diagram.

Signal-dependent capacitance variation can be introduced into Chua's circuit by replacing C_1 and C_2 with the signal-dependent capacitance model characterised by (6.33). The capacitors C_1 and C_2 are replaced by C_1' and C_2', where

$$\begin{aligned} C_1' &= C_1 - \delta C_1 v_{C_1}(t) \\ C_2' &= C_2 - \delta C_2 v_{C_2}(t) \end{aligned} \tag{6.45}$$

Figure 6.15(h)-(i) shows the bifurcation diagrams obtained by plotting the value of $v_{C_1}(t)$ whenever $\dot{v}_{C_1}(t) = 0$ against δC_1 and δC_2 respectively, for the parameter values specified in (6.43). The (dotted) horizontal line for $\delta C_1 > 0.05$ and for $\delta C_2 > 0.2$ in figure 6.15(h) and figure 6.15(i) respectively indicates that the circuit behaviour is characterised by a fixed point. These two bifurcation diagrams suggest that the strange attractor persists for the following signal-dependent values:

$$\begin{aligned} \delta C_1 &\approx< 0.05 \text{ other parameters fixed at the values in (6.43)} \\ \delta C_2 &\approx< 0.2 \text{ other parameters fixed at the values in (6.43)} \end{aligned} \tag{6.46}$$

Matsumoto *et al.* (1985) find that Chua's circuit exhibits deterministic chaos for the range of capacitance (i.e. C_1 and C_2) values specified in (6.44). However, if signal-dependent capacitance variations are made large enough the operation of Chua's circuit is inhibited. This probably occurs because the signal $v_{C_1}(t)$ over part of its cycle takes the capacitance out of the range for chaotic behaviour to occur, and if the dependence (i.e. δC_1 and/or δC_2) is made large enough chaotic behaviour is inhibited.

The sensitivity of the parameters characterising the nonlinearity are assessed by the bifurcation diagrams shown in figure 6.15(f)-(g). These two bifurcation diagrams suggest that the strange attractor persists for the following nonlinearity parameter values:

$$\begin{aligned} -0.4 \approx< m_0 \approx< -0.6 &\text{ other parameters fixed at the values in (6.43)} \\ -0.5 \approx< m_1 \approx< -3.0 &\text{ other parameters fixed at the values in (6.43)} \end{aligned} \tag{6.47}$$

Charge carrier transit delay through the nonlinearity (6.42) causes the current flowing through the nonlinearity to be delayed by $t_r(t)$ with respect to the applied voltage $v_{C_1}(t)$ (i.e. $i(t) = f(v_{C_1}(t - t_r(t)))$). The bifurcation diagram shown in figure 6.15(e) highlights the qualitative behaviour of Chua's circuit when signal delay is introduced into the nonlinearity. Figure 6.15(e) shows the diagram obtained by plotting the value of $v_{C_1}(t)$ whenever $\dot{v}_{C_1}(t) = 0$ against t_d, for the parameter values specified in (6.42). Although the fundamental period of $v_{C_1}(t)$ is approximately 3 sec (*cf.* figure 6.15(a)) a signal delay greater than ≈ 0.01 seconds (i.e. 0.3% of the the fundamental period) is enough to inhibit the operation of Chua's circuit (i.e. the strange attractor is destroyed and a fixed point attractor is formed). Regular (i.e. nonchaotic) behaviour is re-established for signal delays between ≈ 0.5 to 2.5, which is the range of delays plotted on the bifurcation diagram shown in figure 6.15(e).

Chua *et al.* (1986) show that it is possible to systematically generate an infinite number of circuits (termed Chua's circuit family) with different topologies but which

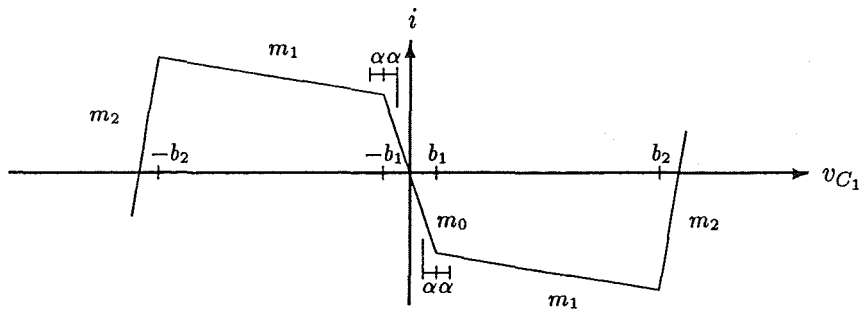


Figure 6.16: Diagrammatic representation of the 5-segment piecewise-linear function characterising the nonlinear resistor in Chua's circuit. The nonlinear function is specified by the slopes m_0, m_1 and m_2 and the corners b_1 and b_2. The intervals labelled α indicate the position where two cubics are smoothly fitted to the piecewise-linear function.

are equivalent to Chua's circuit in the sense that they have identical qualitative behaviour (refer to §2.3). The circuit characterised by (6.41) and (6.42) represents one realization that has been studied in the literature (*cf.* Chua *et al.* 1987). It is nevertheless just one possible realisation from the family (*cf.* Wu 1987, page 1031). By suitably choosing C_1, C_2, G and L it is possible to generate the same qualitative behaviour for a wide range of different 5-segment piecewise-linear functions. The qualitative behaviour of the circuit realization studied in this section seems to be insensitive to many of the circuit parameters. However, computational evidence suggests that the chaotic behaviour generated by Chua's circuit is quite sensitive to:

- The introduction of a signal delay into the nonlinearity. A delay of greater than 0.01 (i.e. greater than 0.3% of the fundamental period of the oscillator signal) reduces the behaviour to a fixed point (provided also the the signal delay is less than 0.5). In §6.4.3 signal delay and particular nonlinearity shapes are shown to be sufficient conditions for the nonlinear oscillators studied in §6.4 to exhibit deterministic chaos. In the case of Chua's circuit it is the absence of signal delay which is necessary. This suggests that deterministic chaos is sensitive to the shape of nonlinearities and the presence or absence of signal delay (i.e. if chaos occurs in a system with signal delay then in general it is not chaotic when the signal delay is removed and vice versa).

- The smoothing out of the corners of the 5-segment piecewise-linear function $f(v_{C_1})$ at $v_{C_1} = \pm b_1$ by the smooth fitting of two cubics (i.e. fit cubics to make the first derivative of $f(v_{C_1})$ continous at $v_{C_1} = \pm b_1$). The two cubics are fitted smoothly to the nonlinearity and they extend over the intervals $[b_1 - \alpha,\ b_1 + \alpha]$ and $[-b_1 - \alpha,\ -b_1 + \alpha]$ on the nonlinearity respectively (refer to figure 6.16). Computer simulation shows that for $\alpha = 0.1$ the strange attractor in Chua's circuit is destroyed and is replaced by a fixed point. Although Matsumoto *et al.* (1985 page 804) states that the strange attractor persists if the 5-segment piecewise-linear function is completely replaced by a cubic, certain kinds of small perturbations to the 5-segment piecewise-linear function reduce the behaviour to a fixed point.

- The slope of the middle segment m_0 of the 5-segment piecewise-linear function. The strange attractor persists when m_0 is in the interval -0.4 to -0.6. Outside

this interval the behaviour reduces to a fixed point.

6.7 Computational Considerations

The differential equations characterising the oscillator models presented in this Chapter have been solved using a numerical method package called Stride (Butcher *et al.* 1979), which implements an implicit Runge-Kutta method (*cf.* Atkinson 1978, §6.9) and was chosen for the following reasons:

- It was readily available.
- Extensive convergence and stability information can be outputted at user specified time intervals.
- Its execution time is similar to that of other methods (e.g. linear multistep methods) which I assessed. The differential equations characterising Chua's circuit (6.41) were used as a benchmark for comparing and evaluating the other software packages tried.
- Runge-Kutta methods are known to process good stability and high accuracy (Atkinson 1978).

Calculation of a numerical solution to a differential equation, which incorporates a delay, requires storage of sequences of past values of state variables. The value of a state variable (e.g. $v(t)$) at some previous instant (e.g. $t - t_r(t)$) is obtained by interpolating the previously stored state variables to obtain the value of the state variable at the required instant. I used a third order Newton interpolation (*cf.* Atkinson 1985, Chapter 3). To verify the accuracy of my numerical results, I solved a time-delay differential equation characterising blood production developed by MacKay and Glass (1977), and compared my results with the extensive numerical investigation of this equation undertaken by Farmer (1982). In particular, I was able to reproduce figures 1,2 and 3 of Farmer (1982) effectively identically.

Two types of error are responsible for any difference between the numerically calculated and the true solution of a differential equation. They result from the replacement of the differential equation by a finite difference equation, and from computational error due to the employment of finite precision numbers (*cf.* Atkinson 1978, §6.1). The first type of error is minimised by reducing the step size (i.e. the intervals between instants for which calculations are made). However, for a chaotic solution, changing the step size leads to a new realization, and the difference between realizations calculated with different intervals can be appreciable even over quite short time intervals. This effectively prevents any numerical method from accurately calculating a chaotic trajectory beyond a certain time interval (*cf.* Lichtenberg and Lieberman 1983, §5.2d; Belyayev *et al.* 1985). However, numerical solutions can demonstrate that particular system equations are likely to be chaotic by demonstrating that there are trajectories which do not appear to repeat and which are also sensitively dependent on initial conditions. The curve in figure 6.10(i) demonstrates that (6.35) is likely to be chaotic for $g > 4$ since the solution does not appear to repeat and it shows apparent sensitivity to initial conditions.

Chapter 7

Chaotic Data Encryption

This Chapter indicates how deterministic chaos might be invoked with advantage for *cryptology*, which is the accepted term for the science of secret communications. The word cryptology is formed from two Greek words meaning 'hidden' and 'word' (Simmons 1985). The aim of cryptology is to encrypt or encipher messages generated by a source, here called *transmitter*, in such a way as to make the contents of the messages secret to everyone except certain specified people, here called *receivers* (Meyer and Matyas 1982; Beker and Piper 1982; Simmons 1988). Cryptology is usually split into two branches: *cryptography* and *cryptanalysis* (Massey 1988). Cryptography is the study of encryption, and encompasses methods for ensuring the secrecy and authenticity of messages (Massey 1988). Cryptanalysis is the study of ways of breaking an encryption or forging coded signals so that they are accepted as authentic (Brickell and Odlyzko 1988).

Although the specific use of chaotic dynamical systems for message encryption seems to have been overlooked, Wolfram (1985) has suggested that cellular automata (a field related to deterministic chaos; refer to Wolfram (1984) and the seventh to last paragraph in Chapter 4) may be useful for message encryption. Wolfram's suggestion is, however, very different from the approach introduced in this Chapter. Chaotic dynamical systems have properties (refer to §1.1.8) which appear to be highly desirable for message encryption. The most important of these properties is that points in state space which are initially close together separate widely as time proceeds. This implies that, if a sequence of seemingly random numbers (i.e. a trajectory) is to be predicted, at least one number in the sequence must be specified exactly. This Chapter discusses to what extent the seemingly random numbers generated by chaotic dynamical systems are suitable for data encryption. Message encryption via chaotic dynamical systems is here called *chaotic encryption.*

It is appropriate here to mention the method for encrypting data messages known as the *one time tape* (Meyer and Matyas 1982, page 20). The method involves combining the message to be encrypted with a key. If the key is made equal to or made longer than the length of the message, then in principle an unbreakable encryption can be achieved (Massey 1988). Unfortunately, this also makes the one time tape impractical for most applications where there is considerable message traffic, since a large number of keys must be transported and stored before communications can be established. It is explained in §7.1 how a chaotic dynamical system can be used to modify the one time tape to render it potentially more practicable. In order

to make the argument developed in §7.1 understandable, cryptologic terminology is introduced and explained at the beginning of §7.1.

Computers can only represent a discrete subset of the real numbers (Atkinson 1978, page 12). A continuous interval of real numbers cannot therefore be represented. When a continuous chaotic dynamical system of equations is modelled on a computer, the dynamics are no longer continuous, but are unavoidably quantised, assuming values belonging to a discrete set of real numbers (McCauley 1988, page 50). The consequences of this are discussed in §7.2.

It is convenient to split the methods for encrypting data messages using chaotic dynamical systems into two categories. A chaotic dynamical system of equations, acting as a random number source, serves as the basis for each category. The first category, here termed *isolated chaotic encryption*, comprises those methods for which the dynamics of the chaotic system are unaffected by the message to be encrypted (i.e. the chaotic system operates in isolation). For methods belonging to the second category, the messages to be encrypted alters the dynamics of the chaotic system. This second category is here called *influenced chaotic encryption*. Isolated and influenced chaotic encryption are discussed in §7.3 and §7.4 respectively. The hardware implementation of chaotic encryption is studied in §7.5.

New results are presented in §7.2, §7.3 and §7.4. The conclusions that can be drawn from the material presented in this Chapter, together with suggestions for further work, are included in Chapter 9.

7.1 Cryptology

The message to be encrypted is called the plaintext message or *plaintext*. The encrypted message is called the *cryptogram*. Encrypting always employs a *key*, which it is hoped is permanently denied to anybody (called hereafter an *enemy*) not specifically authorised to know it (Meyer and Matyas 1982, page 1). Breaking an encryption involves the enemy determining this secret key. An attempt to break an encryption is called an *attack*. An attack on an encryption, when both the cryptogram and the encryption algorithm are known, is called a *cryptogram-only attack*. If in addition some plaintext-cryptogram pairs formed with the actual secret key are available, the attack is called a *known-plaintext attack*. The enemy may have access to the encryption equipment containing the secret key, and be able (depending on the particular implementation) to encrypt selected plaintext (such an attack is called a *chosen-plaintext attack*) or decipher selected cryptogram (such an attack is called a *chosen-cryptogram attack*). This does not amount to the same thing as breaking the encryption, since access to the encryption equipment, if possible at all, is usually very limited, and the secret key remains unknown to the enemy. Except in rare cases, is it possible to state absolutely that certain access and information will never become available to the enemy. Therefore, a conservative approach must be taken to encryption algorithm design. The *universal assumption* of cryptology is that the enemy has full access to the cryptogram, and full knowledge of the encryption method (Massey 1988). This implies that the security of the encryption must reside entirely in the secret key. Most encryption algorithms in use today are intended by their designers to be secure against at least a chosen-plaintext attack (Massey 1988).

Recalling the notation introduced in §2.1, an arbitrary sequence of numbers is

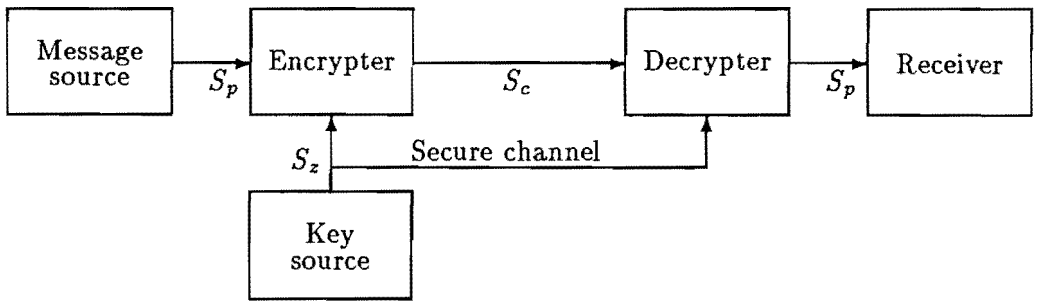


Figure 7.1: An archetypal block-diagram of a secret key cryptosystem. The message source generates the plaintext $S_p = \{p_i : i = 1, \cdots, m\}$. The encrypter forms the cryptogram $S_c = \{c_i : i = 1, \cdots, n\}$ from the secret key and the plaintext.

denoted by $S_x = \{x_i : i = 1, \cdots\}$, where x_i is the i^{th} number in the sequence S_x. A particular S_x is denoted $S_x(x)$.

An encryption system, employing a single secret key hopefully known only to the transmitter and the receivers, is called a *secret key cryptosystem*, an archetypal block-diagram of which is shown in figure 7.1 (Massey 1988; Longe 1983; Diffe and Hellman 1979). The secret key $S_z = \{z_i : i = 1, \cdots, k\}$ consists of k digits. The key is delivered to the intended transmitter and receivers by a *secure channel* (Massey 1988), which is a communication connection which is hopefully very difficult for the enemy to intercept (e.g. a secure channel could be a locked brief case chained to a person's arm). The k digits of the key are chosen from a finite alphabet, which is usually the same as that in which the plaintext is represented. In practice, plaintexts are usually stored in computers, and are therefore represented by sequences of binary digits. This Chapter only considers the encryption of messages which are represented by sequences of binary digits (i.e. the plaintext alphabet consists of a set of two digits $\{0, 1\}$) (Meyer and Matyas 1982, page 23). However, generalisation to an L-ary alphabet is straightforward. The message source generates the plaintext $S_p = \{p_i : i = 1, \cdots, m\}$. The encrypter forms the cryptogram $S_c = \{c_i : i = 1, \cdots, n\}$ which is a function of both the plaintext S_p, and the secret key S_z i.e.

$$S_c = f(S_p, S_z) \tag{7.1}$$

where $f(\cdot)$ is termed the encryption function. The decrypter is the inversion of the encryption function

$$S_p = g(S_c, S_z) \tag{7.2}$$

where $g(\cdot)$ is termed the decryption function.

There are two different notions of security, known as *theoretical* or *unconditional security* and *practical* or *conditional security* (Meyer and Matyas 1982, page 607). If an encryption cannot be broken (i.e. the key cannot be determined) given unlimited time, manpower and number of cryptograms, then the encryption is said to be unconditionally secure, otherwise it is conditionally secure. Most cryptosystems in use today are conditionally secure. For these cryptosystems, the known algorithms for breaking the encryption requires so much effort that for all practical purposes it is not feasible. Shannon (1948) defined a concept called *perfect secrecy* which is a sufficient condition (i.e. a stronger condition than necessary) for unconditional security.

An encryption achieves perfect secrecy if the plaintext S_p is statistically independent of the cryptogram S_c (i.e. $P(S_p = S_p(p) \mid S_c = S_c(c)) = P(S_p = S_p(p))$ for all possible plaintexts $S_p(p)$ and cryptograms $S_c(c)$) (Massey 1988). This means that enemy cryptanalysts can do no better estimating S_p with knowledge of S_c than they could do without knowing S_c, no matter how much time and manpower they have available. Only some cryptosystems are known to possess perfect secrecy (Massey 1988).

Shannon (1948) has shown that a cryptosystem can only possess perfect secrecy if the plaintext is completely random (i.e. does not contain redundancy) and the secret key is at least as long as the plaintext. The only perfectly secret system in common use is the one time tape (Massey 1988). Its encryption function is

$$c_i = p_i \oplus z_i, \quad i = 1, 2, \cdots, M \tag{7.3}$$

where the plaintext, cryptogram and key digits belong to a L-ary alphabet $\{0, 1, \cdots, L-1\}$ and $\oplus$ denotes addition modulo L (for the digital implementations considered in this thesis L is understood to be limited to two). Since $P(S_c = S_c(c) \mid S_p = S_p(p)) = L^{-M}$ for every possible $S_c(c)$ and $S_p(p)$, then no matter what the statistics of $S_p(p)$ are, S_p and S_c are statistically independent, which implies perfect secrecy (Massey 1988). The one time tape requires at least one digit of secret key for each digit of plaintext. This is impractical for most applications (i.e. applications where there is a reasonable flow of plaintext digits to be encrypted) since this results in the need for very long keys. The generation, distribution and the storage of keys at the transmitter and receivers becomes increasingly impractical as the key becomes very long. The generation and distribution becomes time consuming and difficult to manage, and the secure storage of such keys at the transmitter and receivers becomes increasingly difficult. However, it is possible to use instead a pseudo-random number generator to generate the long keys which are needed to implement the one time tape. Such a cryptosystem is then termed a *stream cipher* (Meyer and Matyas 1982, page 53), and the sequence of numbers generated by the pseudo-random number generator is termed the *key stream*. A secret key must still be distributed to the transmitter and receivers, but is used as a seed for the generator. The secret key is much shorter than the key stream generated by the pseudo-random number generator. For it to possess perfect secrecy, despite the shortness of the key, the pseudo-random number generator must possess the properties:

- All numbers in the generated sequence must be statistically independent of each other (i.e. every possible finite length sequence of numbers must occur with equal probability).
- It must not be possible to deduce the entire sequence from a small segment of the sequence.
- The number generator must be deterministic because both the transmitter and all receivers must be able to generate identical sequences.
- Each seed for the number generator must generate a different sequence.

Most pseudo-random number generators which have been proposed and implemented in computer programs and/or directly in hardware do not have all these properties.

For instance, pseudo-random sequences generated by an n-stage shift register with feedback, can be entirely deduced by observing $n(n-1)$ terms in the sequence (Diffe and Hellman 1979).

A practical implementation of chaotic encryption requires a digital computer (e.g. microprocessor) and/or dedicated hardware. Such an implementation is studied in §7.5. All digital computers represent real numbers to a finite precision (i.e. by a finite number of binary digits (bits)) (Atkinson 1978, page 12). This has two consequences. The first is that only a finite number of real numbers can be represented by a particular computer. The second consequence is that there is a real number (called the machine epsilon), which is the smallest number that can be represented by a computer. The difference between any two different numbers represented by a digital computer is usually a multiple of the machine epsilon. A computer can therefore only represent a discrete subset of the real numbers. If a process using infinite precision numbers is modelled on a computer, then the resulting model unavoidably uses finite precision numbers. This phenomenon is here called *number discretization* (also called number quantisation or coarse graining). The next section discusses how number discretization affects the dynamics of chaotic dynamical systems.

7.2 The Effect of Number Discretization on Chaos

Any chaotic dynamical system of equations cannot be truly chaotic when implemented on a digital computer. Since a computer can only represent a finite set of real numbers, the longest (i.e. the maximal length) sequence of numbers that can be generated without repetition is necessarily finite. This implies that a sequence of numbers generated from a chaotic dynamical system of equations must, when implemented on a computer, repeat after a finite number of iterations, and hence cannot be genuinely chaotic. This section attempts to assess the effect of the computer's finite precision on the dynamics of a chaotic dynamical system of equations. The most suitable chaotic dynamical systems for pseudo-random number generation are those that generate maximal length (seemingly random) sequences when restricted to finite precision numbers. However, it is not immediately obvious how to synthesis such chaotic systems. The approach taken in this section is to observe through computer simulations how a pair of different chaotic dynamical systems are affected by the finite precision. Number discretization alters the dynamics of the chaotic system of equations. As regards data encryption, the important aspects of the dynamics of a chaotic system of equations are:

- The length p of the period of the sequence (i.e. the number of times that the chaotic equations must be iterated before the sequence repeats).
- The amount of information the sequence contains (i.e. the entropy of the sequence).

This section studies the first aspect, while §7.3 and §7.4 study the second aspect.

The length of the period of a sequence is affected by the size of the number quantisation step (i.e. the machine epsilon). An indication of the effect of number discretization can be assessed by observing how quickly or how soon (i.e. after how many iterations) a trajectory returns close to any particular point on the trajectory

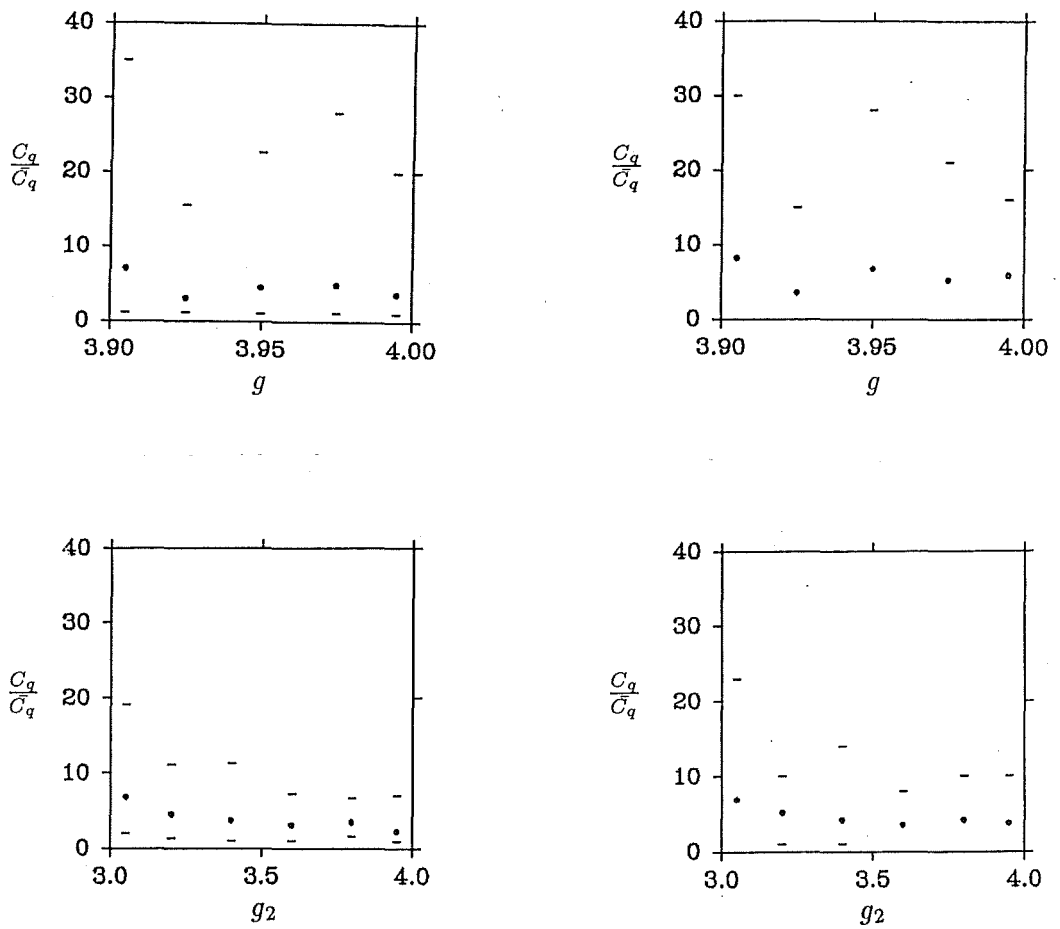


Figure 7.2: Assessment of the uniformity property. The top and bottom pair of graphs show respectively the results obtained from iterating the canonical equation for values of g lying between 3.9 and 4.0, and from iterating a hierarchy of two canonical equations for $g_1 = 3.9$ and for values of g_2 lying between 3.0 and 4.0. The graphs in the left and right hand column show results for $q = 10^{-3}$ and $q = 10^{-5}$ respectively. The vertical axis represents the normalized C_q values (i.e. $C_q/\bar{C}_q$). The small circles represent the mean value of $C_q/\bar{C}_q$, and the short horizontal lines represent the maximum and minimum value.

(i.e. how quickly a trajectory returns close to itself). Consider a chaotic system where the trajectory returns within the machine epsilon of a particular point in a relatively short time (i.e. after only a few iterations). When such a chaotic system is iterated using quantised numbers, it is to be expected that the sequence of numbers generated from the system must repeat within a short period. Therefore, discrete chaotic systems whose trajectories are such that distances between points on them are maximised, would appear to hold the most promise. Such systems are here said to possess the 'uniformity' property. The only practical way of assessing to what degree a particular chaotic equation possesses this property seems to be to perform computer simulations. The results of two such simulations, one on the canonical equation (i.e. logistic map; refer to §1.1.6 and first paragraph of Chapter 5) and one on a hierarchy of two canonical equations (refer to §5.4), are presented in this section. As a reminder, the canonical equation is characterised by the recursive equation

$$x_{n+1} = g x_n (1 - x_x) \tag{7.4}$$

where g is a parameter, and a hierarchy of two canonical equations is characterised by the recursive equation

$$\begin{aligned} y_{n+1} &= g_1 y_n (1 - y_n) \\ x_{n+1} &= g_2 x_n y_n (1 - x_n) \end{aligned} \tag{7.5}$$

where g_1 and g_2 are parameters, and where the sequence $\{x_i : i = 0, 1, \cdots\}$ is taken as the sequence generated by the hierarchy.

The computer simulations performed on (7.4) and (7.5) uses 16 decimal digit numbers and consists of the following steps:

1. Choose an initial condition pseudo-randomly (using whatever random function routine is standard for the available computer - VAX Fortran for the simulations reported herein).

2. Iterate the chaotic dynamical system m times, and check if the chosen initial condition leads to a seemingly chaotic sequence of numbers. Return to step 1 if the chosen initial condition does not.

3. Iterate another n times, measuring the Euclidean distance d_i between x_m and x_{m+i} for $i = 1, \cdots n$ (i.e. $d_i = \mid x_m - x_{m+i} \mid$ for $i = 1, \cdots n$).

4. For every $d_i < q$ (where q is the size of the quantisation step), increment C_q (C_q is initially set to zero).

5. Return to step 1, and repeat the procedure 100 times.

The expected value of C_q for an equation having the uniformity property is $\bar{C}_q = nq/I$, where I is the domain of the equation ($I = 1$ for both the canonical equation and the hierarchy). The top and bottom pair of graphs in figure 7.2 show respectively the results obtained from iterating the canonical equation for values of g lying between 3.9 and 4.0, and from iterating a hierarchy of two canonical equations for $g_1 = 3.9$ and for values of g_2 lying between 3.0 and 4.0. The small circles represent the mean value of $C_q/\bar{C}_q$, and the short horizontal lines above and below the mean value represent the maximum and minimum value respectively. The value of the

gain parameter for which the ratio $C_q/\bar{C}_q$ plotted in figure 7.2 is a minimum gives the value for which the uniformity property is optimal. Since at each value of gain the ratio $C_q/\bar{C}_q$ is characterised by a pdf, the minimum of the mean values of $C_q/\bar{C}_q$ plotted in figure 7.2 are taken to specify the gain value for which the uniformity property is optimal. Inspection of the graphs in figure 7.2 show that the value of the gain parameter for which the canonical equation and the hierarchy possess optimal unformity is when $g = 4.0$ and $g_2 = 4.0$ respectively.

In order to study, by computer simulation, the effect of number quantisation on the length of the period of a sequence of numbers generated by a particular chaotic dynamical system, the following steps are implemented:

1. Choose a discrete set of uniformly spaced numbers N_q, extending over the domain I and spaced by the quantisation step q.

2. Choose an initial condition pseudo-randomly (as in step 1 of the simulation previously discussed in this section) from the set N_q.

3. Iterate the chaotic dynamical system m times, and check if the initial condition leads to a seemingly chaotic sequence of numbers. Return to step 1 if the chosen initial condition does not.

4. Continue iterating until the sequence repeats. Each iteration x_i is truncated by choosing from the set N_q the number closest to x_i which is also less then x_i.

5. Return to step 2, and repeat the procedure 100 times.

In this simulation the parameters characterising the canonical equation and the hierarchy, are set to the values for which the uniformity property is optimal (i.e. $g = 4.0$, $g_1 = 3.9$, $g_2 = 4.0$). The top and bottom pair of graphs in figure 7.3 show the results obtained from iterating the canonical equation and the hierarchy respectively. The graphs in the left hand column of figure 7.3 give the maximum, minimum and mean period length versus the quantisation step q. The mean period length is represented by a circle, and the maximum and minimum is represented by short horizontal lines. The left hand column of graphs show that the period of the sequence rapidly increases with reducing q, for values of q below about 10^{-9}. The computational expense required for computing the results in figure 7.3, precludes obtaining results for q below about 10^{-11}. The graphs in the right hand column of figure 7.3 give the maximum, minimum and mean normalized period pq (where p denotes the length of the period of the sequence) verses the quantisation step q. The right hand column of graphs show that for q below about 10^{-9}, p increases at a faster rate (approximately two times faster) with q, than for q above 10^{-9}. This is of practical significance because it shows that greater gains in p can be achieved for q smaller than 10^{-9}, than for q larger than 10^{-9}. A q of 10^{-16} can easily be obtained with existing hardware technology. Extrapolating the points in the right hand column of the graphs in figure 7.3 suggests $pq \approx 10^{-4}$ for $q = 10^{-16}$, giving $p = 10^{12}$ which is of considerable length.

The effect of number quantisation on the power spectrum is now studied. Recalling the notation used in the second paragraph of Chapter 5, the power spectrum formed by averaging $(X_i(x_0, N))^2$ over M (pseudo-)randomly chosen values of x_0,

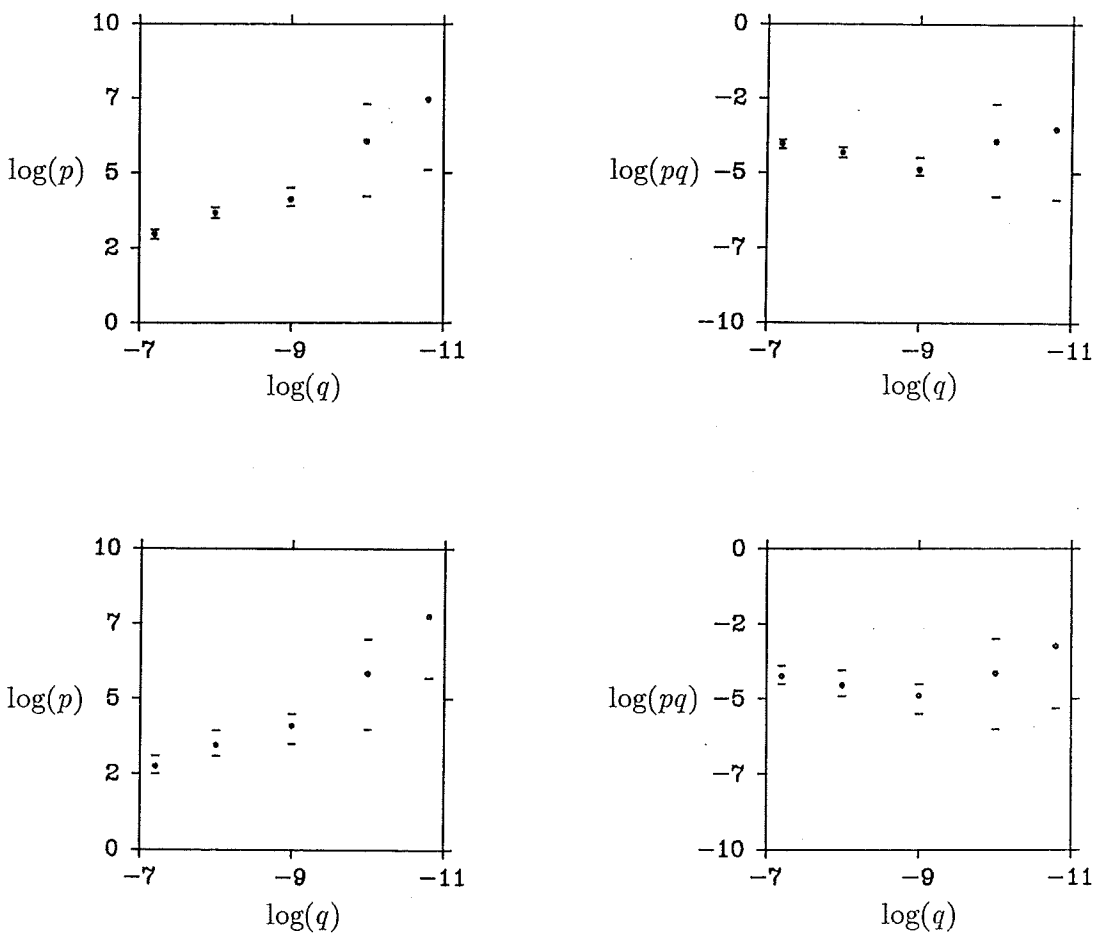


Figure 7.3: The length of the period of the generated sequences. The top and bottom pair of graphs show respectively the results obtained from iterating the canonical equation and the hierarchy. The graphs in the left hand column give the maximum, minimum and mean period length versus the quantisation step q. The graphs in the right hand column give the maximum, minimum and mean normalized period pq verses the quantisation step q. The mean period is plotted as small circles, and the short horizontal lines represent the maximum and minimum value.

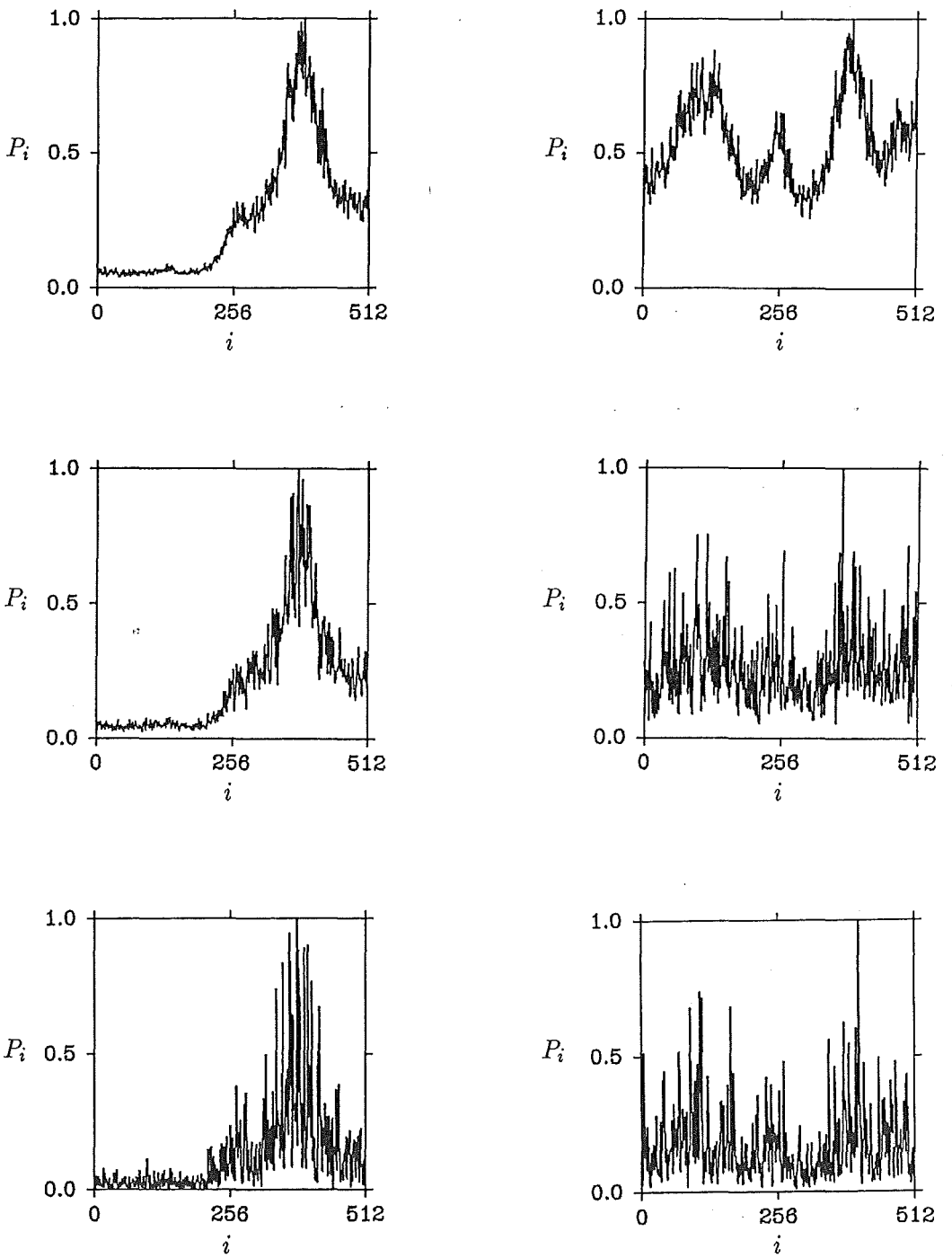


Figure 7.4: The effect of number quantisation on the power spectrum. The power spectra in the left and right hand columns are calculated from sequences generated from the canonical equation and the hierarchy of two canonical equations respectively. The parameter values used in the equations are set to where the uniformity property is optimal (i.e. $g = 4.0$, $g_1 = 3.9$, $g_2 = 4.0$). The top, middle and bottom rows of power spectra are for $q = 10^{-16}$, $q = 10^{-9}$ and $q = 10^{-7}$ respectively.

the m^{th} of which is written as $x_{0,m}$, is denoted by

$$P_i(M,N) = (1/M)\sum_{m=1}^{M}(X_i(x_{0,m},N))^2 \tag{7.6}$$

Figure 7.4 shows six power spectra, for which $N = 1024$ and $M = 100$. The power spectra in the left and right hand columns are calculated from sequences generated respectively form the canonical equation and the hierarchy. The parameter values used in the equations are set to where the uniformity property is optimal (i.e. $g = 4.0$, $g_1 = 3.9$, $g_2 = 4.0$). The top, middle and bottom row of power spectra in figure 7.4 are obtained for $q = 10^{-16}$, $q = 10^{-9}$ and $q = 10^{-7}$ respectively. Figure 7.4 shows that the power spectrum is not affected significantly by the number of significant digits in the finite precision numbers used to generate the sequence. This suggests that the statistics of the sequences are not significantly affected by number discretization. This is of practical importance, since it suggests that the large body of dynamical results obtained for infinite precision numbers also applies to the dynamics generated by using finite precision numbers.

7.3 Isolated Chaotic Encryption

The cryptosystem studied in this section is a stream cipher (refer to §7.1), where the heart of the pseudo-random number generator is a chaotic dynamical system. The block diagram of the chaotic pseudo-random number generator is shown in figure 7.5, where the arrows give the direction of data flow. Block 1 represents a discrete chaotic dynamical system and is the source of apparent randomness for the encryption. The output from this block is a sequence of scalars (or more generally vectors) denoted $\{x_i : i = 1, \cdots\}$. Block 2 represents a transformation T_r, which transforms x_n into r_n (i.e. $T_r : x_n \rightarrow r_n$, where r_n is in general a scalar). The point of this is that it transforms each x_n into a form more suitable for processing into a sequence of binary bits. T_r may involve only scaling or it may be more complicated (e.g. if x_n is a vector, T_r transforms x_n into a scalar). Block 3 represents a transformation T_o, which transforms r_n into r_n^* (i.e. $T_o : r_n \rightarrow r_n^*$). The purpose of T_o is to add in further complication, hopefully making an enemy attack (as outlined in the first paragraph of §7.1) less successful. It is a transformation that is designed not to have a unique inverse. Block 4 represents a transformation T_c, which transforms r_n^* into b_n^* (i.e. $T_c : r_n^* \rightarrow b_n^*$). The purpose of this transformation is to reduce any correlation between consecutive numbers in the sequence (i.e. make the numbers b_n^* statistically independent). Finally, block 5 represents a transformation T_b, which transforms b_n^* into b_n (i.e. $T_b : b_n^* \rightarrow b_n$). This transformation performs a thresholding operation on b_n^* to form a binary digit. The binary digits outputed from block 5 form the key stream for the stream cipher. Each of these blocks (shown in figure 7.5) is discussed in detail below.

The blocks numbered 2 and 4 in figure 7.5 are not always needed. Sometimes the numbers generated by the chaotic dynamical system (i.e. block 1) are already suitably scaled, etc., and are effectively statistically independent. Alternatively, a single transformation may perform the equivalent of several blocks. In such cases only two or three of the four transformations may need to be implemented. For example, blocks 2 and 4 may not be required for the canonical equation with $g = 4.0$,

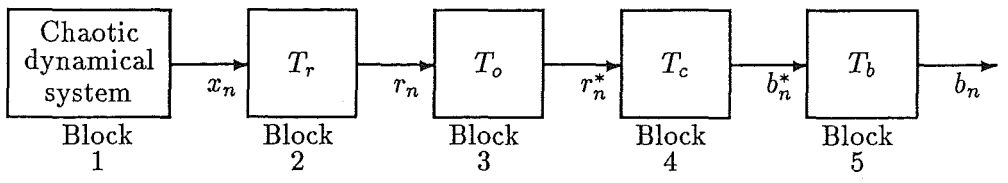


Figure 7.5: Block diagram of the chaotic pseudo-random number generator. The tail of each arrow indicates data flow from a block, the head of each arrow indicates data flow into a block.

since the numbers generated are suitably scaled, and may be sufficiently statistically independent for some encryption applications. A discrete chaotic dynamical system can be written in the form

$$y_{n+1} = f(y_n) \tag{7.7}$$

where y_{n+1}, y_n and $f(\cdot)$ are in general vectors. The calculation of y_{n+1} from y_n is here termed an iteration in forward time or a *forward iteration.* The calculation of y_n from y_{n+1} is here termed an iteration in reverse time or a *reverse iteration.* A one-dimensional chaotic dynamical system exhibits chaotic dynamics (refer to §1.1.8) when iterated in forward time only. If iterated in reverse time, a one-dimensional chaotic system does not exhibit chaotic dynamics (i.e. intervals contract on average with each reverse iteration). The forward and reverse sequence of numbers is the sequence of numbers obtained through iterating (7.7) in forward and reverse time respectively.

If an enemy obtains the value of any y_n then, at the very least, the entire forward sequence can be calculated using (7.7) (it is assumed the enemy has full knowledge of the method of encipherment, as implied by the universal assumption). The value of y_n could be obtained through direct measurement or observation of the encrypting hardware. Fortunately the enemy (hopefully) cannot make direct observations or measurements on the hardware itself. However, an algorithm may exist which might give an estimate of the value of y_n from observation of the cryptogram alone (no such algorithm has yet been developed though). If the value of a sufficient number of members in the sequence $\{y_n : n = 1, 2, \cdots\}$ can be estimated, then it is possible to calculate an initial condition y_m from which at least the forward sequence starting from y_m can be calculated. It is therefore vital that the value of each y_n cannot be determined or even estimated. The possibility of calculating an exact initial condition y_m from a sequence of estimated y_n values is most easily explained with the help of an example. Consider the graph of the canonical equation shown in figure 7.6. The dots on the vertical and horizontal axes represent the actual values of y_{n+1} and y_n respectively. The short line segment on either side of the dots denote the uncertainty interval with which an enemy has managed to estimate y_n and y_{n+1}. A reverse iteration of the uncertainty interval associated with y_{n+1} gives an improved estimate of y_n. If a sufficient number of consecutive estimates of y_n can be obtained, then by reverse iteration of the uncertainty interval the preceding y_n can be estimated more accurately. This process can be continued until a value of y_n is deduced to the required accuracy. This value forms an initial condition from which the entire forward sequence can be calculated.

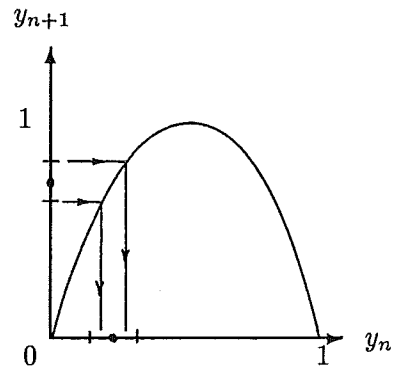


Figure 7.6: The dots on the vertical and horizontal axes give the actual values of y_{n+1} and y_n respectively. The short line segment on either side of the dots denote an uncertainty interval (i.e. the error with which an enemy have estimated y_n and y_{n+1}). A reverse iteration, indicated by the arrows, of the uncertainty interval associated with y_{n+1} gives an improved estimate of y_n.

To eliminate this possible weakness a second dynamical system has been incorporated into block 1 of the chaotic pseudo-random number generator. The second dynamical system exhibits chaotic dynamics when iterated in the reverse direction. It is characterised by

$$z_{n+1} = f^{-1}(z_n) \tag{7.8}$$

where $f^{-1}(\cdot)$ is the inverse function of $f(\cdot)$ in (7.7). The two sequences of numbers, generated by (7.7) and (7.8), are combined to form the output of block 1. The y_n and the z_n are combined in a modulo one adder to form x_n, i.e.

$$x_{n+1} = (x_n + y_n + z_n) \bmod(1) \tag{7.9}$$

where (a) mod(b) returns the remainder of a/b. Since the sum becomes one of the addends in the next addition the operation represented by (7.9) is here termed *cyclic modulo one addition* (*cf.* §7.5). The two dynamical systems (7.7) and (7.8) together with the cyclic modulo one adder (7.9) constitute block 1 in figure 7.5. Even if an enemy develops an algorithm which can estimate x_n, a reverse iteration will not improve the accuracy of preceding estimates. The equation which is chaotic when iterated in reverse time expands the estimated error by the same amount that the equation which is chaotic in forward time contracts the estimated error. It is important that the average rate of error expansion be the same as the average rate of error contraction. More precisely, the Lyapunov exponents (refer to §1.1.12) of the dynamical system, which is chaotic in reverse time, must be the reciprocal of the Lyapunov exponents of the dynamical system which is chaotic in forward time. This is most easily attained if the equations defining the two dynamical systems are each other's inverse. That is, if one is characterised by $x_{n+1} = f(x_n)$, then the other is characterised by $y_{n+1} = f^{-1}(y_n)$, where $f^{-1}(\cdot)$ denotes the inverse function of $f(\cdot)$.

Two different chaotic equations are analysed as possible candidates for the chaotic pseudo-random number generator. The first equation is the canonical equation with $g = 4.0$. The second equation is the hierarchy of two canonical equations with

$g_1 = 3.9$ and $g_2 = 4.0$. The canonical equation does not have a unique inverse. To each value of $y_{n+1} = f(y_n)$ there corresponds two possible values for y_n, one value is always less than 0.5 while the other value is always greater than 0.5. The decision as to which of the two possible values of z_{n+1} to choose at each iteration of the inverse canonical equation is decided by the (forward) canonical equation. If y_n is greater than 0.5 then z_{n+1} is chosen to be greater than 0.5, and vice versa. Thus (7.7) and (7.8) are by necessity usually interlinked. The canonical equation and its inverse are

$$\begin{aligned} y_{n+1} &= g y_n(1-y_n) \\ z_{n+1} &= \begin{cases} \frac{1}{2} + \sqrt{\frac{1}{4} - \frac{z_n}{g}} & \text{if } y_n > \frac{1}{2} \\ \frac{1}{2} - \sqrt{\frac{1}{4} - \frac{z_n}{g}} & \text{if } y_n \leq \frac{1}{2} \end{cases} \end{aligned} \tag{7.10}$$

where $g = 4.0$. The hierarchy of two canonical equations and their inverses are

$$\begin{aligned} u_{n+1} &= g_1 u_n(1-u_n) \\ y_{n+1} &= g_2 u_n y_n(1-y_n) \\ v_{n+1} &= \begin{cases} \frac{1}{2} + \sqrt{|\frac{1}{4} - \frac{v_n}{g_1}|} & \text{if } u_n > \frac{1}{2} \\ \frac{1}{2} - \sqrt{|\frac{1}{4} - \frac{v_n}{g_1}|} & \text{if } u_n \leq \frac{1}{2} \end{cases} \\ z_{n+1} &= \begin{cases} \frac{1}{2} + \sqrt{|\frac{1}{4} - \frac{z_n}{g_2 v_n}|} & \text{if } y_n > \frac{1}{2} \\ \frac{1}{2} - \sqrt{|\frac{1}{4} - \frac{z_n}{g_2 v_n}|} & \text{if } y_n \leq \frac{1}{2} \end{cases} \end{aligned} \tag{7.11}$$

where $g_1 = 3.9$ and $g_2 = 4.0$ and the absolute operation $|\cdot|$ ensures that the square root of a positive number is taken.

A sequence of usable seemingly random scalar numbers $\{r_n : n = 0, 1, \cdots\}$ is obtained by transforming the x_n (the transformation is represented by block 2 in figure 7.5). The purpose of this transformation is to scale x_n and convert it to a scalar if required. Thus

$$r_n = g(x_n) \tag{7.12}$$

where $g(\cdot)$ is the appropriate transformation function. For the case where x_n is already a scalar and no scaling is required, g reduces to the identity $g(x) = x$.

The purpose of T_o (represented by block 3 in figure 7.5) is to add complication, and is formulated to be such that r_n cannot be deduced from the sequence of r_n^*. This requires the inverse of T_o to be many (ideally infinitely) valued. Many potential transformations exist, but two simple ones are:

- Cyclic modulo α addition where α is any real number. For ease of hardware implementation, $\alpha = 1$ is possibly the best choice. The transformation is then of the form $r_{n+1}^* = (r_n^* + r_n) \bmod(1)$
- Moving average equation of the form $r_{n+1}^* = \alpha r_n^* + r_n$, where α is an arbitrary constant less than unity.

The transformation T_o chosen for analyses in this section is the cyclic modulo one addition, i.e.

$$r_{n+1}^* = (r_n^* + r_n) \bmod(1) \tag{7.13}$$

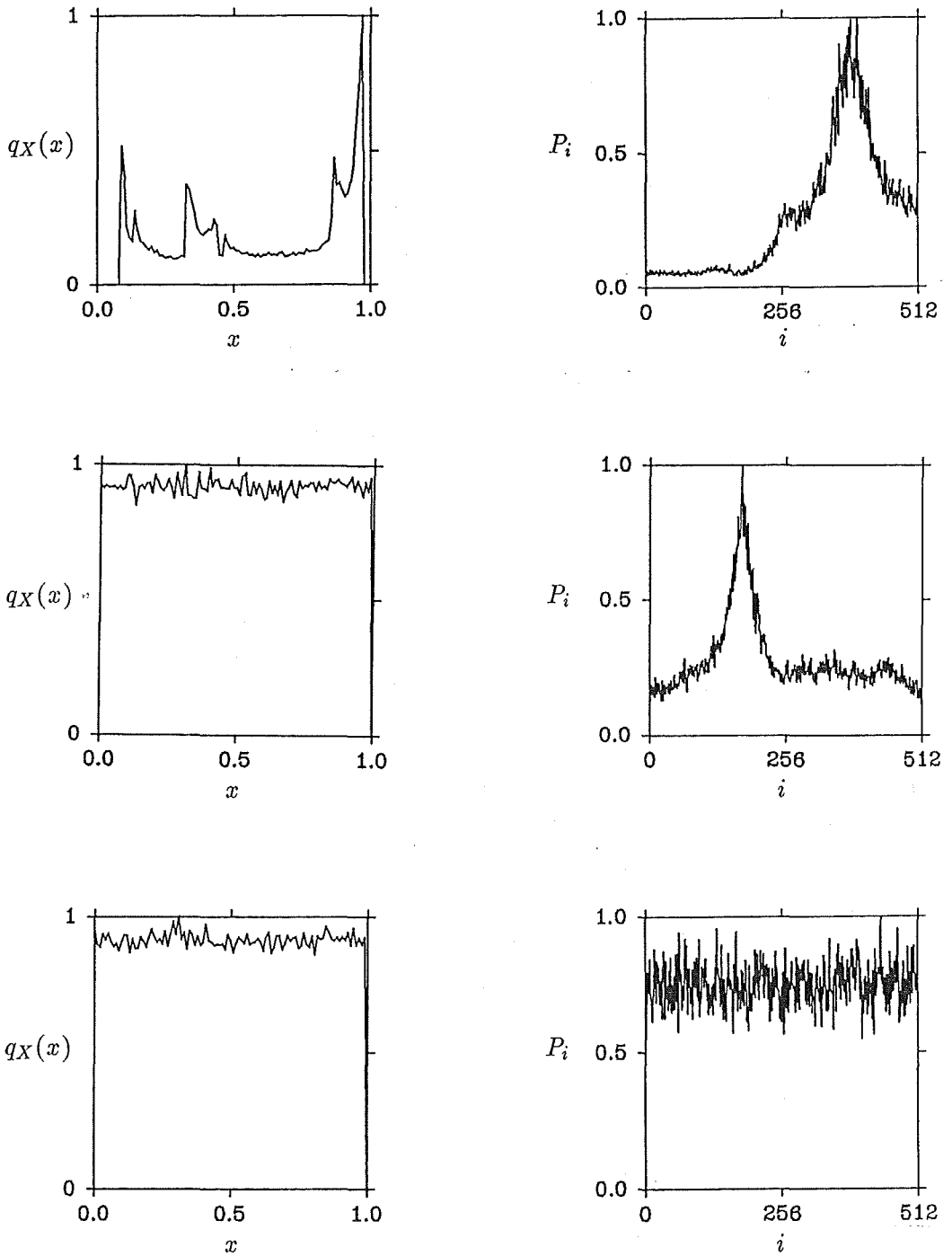


Figure 7.7: The result of cyclic modulo one additions. The top row of curves show respectively the non-uniform pdf and power spectrum of a sequence of numbers generated by a canonical equation for $g = 3.9$. The middle and bottom row of curves show the pdf and PS after one and three (i.e. with $i = 1$ and 3) cyclic modulo one additions respectively.

The binary digits b_n can only be statistically independent if the x_n are statistically independent. The numbers generated by a chaotic system in general have a finite length autocorrelation (Ott 1981, page 656). The pseudo-random numbers x_n can be made effectively independent by a number of methods. One simple method is to use only those members of the sequence $\{x_n : n = 0, 1, \cdots\}$ which are p members apart, where p is made large enough to ensure that the correlation between members p apart is small. Another method of making the numbers of the sequence $\{x_n : n = 0, 1, \cdots\}$ effectively independent is to take p consecutive members of the sequence $\{x_n : n = 0, 1, \cdots\}$ (a p length block) and randomly mix up (i.e. shuffle) the members in the block. Neither method is very elegant, however. Fortunately, while performing computer simulations, a method for transforming sequences of numbers that do not have a uniform (i.e. flat) power spectrum and/or amplitude pdf into sequences of numbers with a uniform power spectrum and amplitude pdf was found. This transformation, which consists of cyclic modulo one adders in series, is

$$\begin{aligned} u_{1,n+1} &= (u_{1,n} + u_{0,n}) \bmod(1) \\ u_{2,n+1} &= (u_{2,n} + u_{1,n}) \bmod(1) \\ &\vdots \\ u_{i,n+1} &= (u_{i,n} + u_{i-1,n}) \bmod(1) \end{aligned} \tag{7.14}$$

where the sequence $\{u_{0,n} : n = 0, 1, \cdots\}$ has a non-uniform power spectrum and pdf, and the sequence $\{u_{i,n} : n = 0, 1, \cdots\}$ has a uniform power spectrum and pdf. Figure 7.7 shows the result of applying cyclic modulo one additions in series. The top row of curves show the non-uniform pdf and power spectrum of a sequence of numbers generated by a canonical equation for $g = 3.9$. The middle and bottom row of curves show the pdf and PS after one and three (i.e. with $i = 1$ and 3) cyclic modulo one additions respectively. After more then three cyclic modulo one additions little improvement in the flattness of the PS or pdf is obtained. Although all curves in the top row of figure 7.7 are markedly non-uniform, all curves in the bottom row are closely uniform. Because, as explained in the previous paragraph, the transformation T_o (represented by block 3) is performed by a cyclic modulo one addition, transformation T_c (represented by block 4) only needs to perform two cyclic modulo one additions in series. Transformations T_o and T_c together give three cyclic modulo one additions in series.

Transformation T_b (represented by block 5) maps each real number b_n^* into a binary digit (i.e. 0 or 1). Transformation T_b is achieved through thresholding the b_n^*. The thresholding is such that it is relatively straightforward to implement in software and/or hardware i.e.

$$b_n = \begin{cases} 0 & \text{if } 5i/10^h < b_n^* \leq 5(i+1)/10^h \text{ for } i = 0, 2, \cdots 2 \times 10^{n-1} - 2 \\ 1 & \text{if } 5i/10^h < b_n^* \leq 5(i+1)/10^h \text{ for } i = 1, 3, \cdots 2 \times 10^{n-1} - 1 \end{cases} \tag{7.15}$$

where h is an integer greater than 0. In this section h is taken to be 10.

Table 7.1 summarises the equations characterising the chaotic pseudo-random number generator shown in figure 7.5.

The greater the amount of information required to specify a sequence the less redundancy it exhibits (i.e. a sequence is more random, in the algorithmic sense, than a sequence which possesses less entropy; refer to §3.1). Entropy is a measure of

Table 7.1: Summary of the equations characterising the chaotic pseudo-random number generator.

Block number	Equation number
1	(7.9) and (7.10) or (7.9) and (7.11)
2	not implemented explicitly
3	not implemented explicitly
4	(7.14)
5	(7.15)

the amount of information a sequence contains. By definition, sequences generated by an ideal pseudo-random number generator possess maximum entropy (note that, maximum entropy implies statistical independence). The entropy of a sequence is equal to the average amount of information contained in each sequence symbol, where such a symbol could be a single member of the sequence or a finite length subsequence. Entropy, denoted H, is defined by

$$H = -\sum_j P_j log_2 P_j \tag{7.16}$$

where P_j is the probability that the source generates the j^{th} symbol. Entropy is maximised when each symbol occurs with equal probability. The entropy of a pseudo-random source is a maximum when every possible arbitrarily long sequence has equal probability of occurring. To determine the information content of the sequences generated by the chaotic pseudo-random number generator two tests which measure entropy are made. The purpose of the first test is to determined if all possible consecutive subsequences of length n, which are obtained from a single generated sequence, occur with equal probability. The purpose of the second test is to determined if all possible sequences of length n, where each sequence is started from a randomly chosen initial condition, occur with equal probability. Exhaustive tests (i.e. for n indefinitely large) are impractical because of the computational expense. Because of this, results are presented for $n = 8$ only. The first test consists of the steps:

1. Choose an initial condition pseudo-randomly (using whatever random function routine is standard for the available computer - VAX Fortran for the simulations reported herein).

2. Use the initial condition chosen in step 1 as a seed for the chaotic pseudo-random number generator. Iterate the number generator 10,000 times (to reduce any initial transients) to generate the output sequence $\{b_i : i = 1, \cdots, 10,000\}$.

3. Generate the output binary sequence $\{b_i : i = 10,001, \cdots\}$.

4. Split $\{b_i : i = 10,001, \cdots\}$ into the m consecutive subsequences $S_b(j) = \{b_i : i = (j-1)n + 10,001, \cdots, jn + 10,000\}$ for $j = 1, \cdots m$.

5. Interpret each $S_b(j)$ as a binary number, where $\{b_i : i = 1\}$ is interpreted as the least significant bit and $\{b_i : i = n\}$ is interpreted as the most significant bit.

6. Plot the interpreted numbers on a histogram where the vertical axis gives the frequency of occurrence, and the horizontal axis lists the ordered numbers $0, \cdots, 2^n - 1$.

This simulation calculates the frequency of occurrence with which consecutive subsequences of length n (obtained from a single generated sequence) occur. For a sequence to be considered random, subsequences of length n must occur with approximately equal probability.

The second test is achieved through the computer simulation consisting of the following steps:

1. Choose m initial conditions pseudo-randomly (as in step 1 of the simulation previously discussed in this section).

2. Use each of the m initial conditions chosen in step 1 as the seeds for m chaotic pseudo-random number generators. Iterate each of the m number generators 10,000 times (to reduce any initial transients) to generate m output sequences $\{b_i : i = 1, \cdots, 10,000\}$.

3. Generate the next n binary bits for each of the m sequences $S_b(j) = \{b_i : i = 10,001, \cdots, 10,000 + n\}$ for $j = 1, \cdots, m$.

4. Interpret each $S_b(j)$ as a binary number, where $\{b_i : i = 1\}$ is interpreted as the least significant bit and $\{b_i : i = n\}$ is interpreted as the most significant bit.

5. Plot the interpreted numbers on a histogram where the vertical axis gives the frequency of occurrence, and the horizontal axis lists the ordered numbers $0, \cdots, 2^n - 1$.

This simulation calculates the frequency of occurrence with which sequences of length n (where each sequence of length n is started from a different randomly chosen initial condition) occur. The sequences generated by an ideal pseudo-random number generator are different for each initial condition, and as a consequence all sequences of length n started from different initial conditions should occur with approximately equal probability.

The left and right hand column of histograms in figure 7.8 show the results obtained from the first and second simulation respectively. The top and middle pair of histograms show the results for the canonical equation (7.10) for $m = 10^5$ and $m = 10^6$ respectively. The wiggles in the middle pair of histograms are smaller than the top pair, indicating that the wiggles are the result of statistical sampling. The bottom pair of histograms show the results for the hierarchy (7.11) for $m = 10^6$. The left hand column of histograms indicates the results for the first simulation and shows that every subsequence of 8 bits occurs with approximately equal probability (and as a consequence all subsequences of less than 8 bits must also occur with equal probability). The right hand column of histograms indicate the results for the

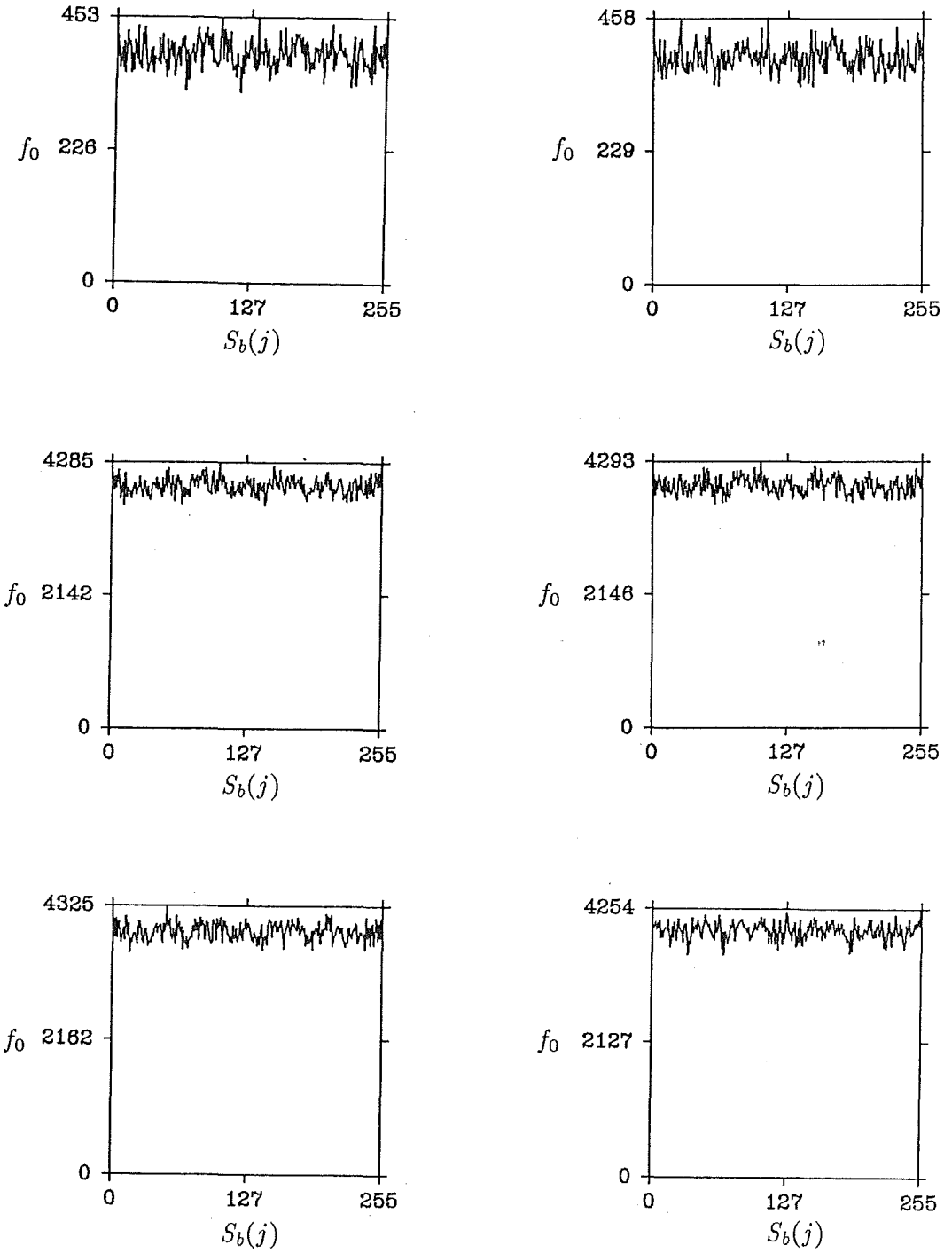


Figure 7.8: Results assessing the extent to which the chaotic pseudo-random number generator reaches the ideal. The left and right hand column of histograms plot respectively the results from the first and second computer simulation. All results are for $n = 8$. The horizontal axis lists the 256 combinations possible with 8 bits. The vertical axis gives the frequency of occurrence f_o with which each 8 bit sequence occurs. The top and middle histograms give the results for the canonical equation (7.10) for $m = 10^5$ and $m = 10^6$ respectively. The bottom row give the results for the hierarchy (7.11) for $m = 10^6$.

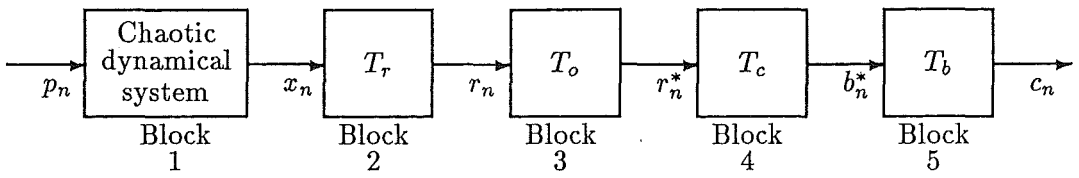


Figure 7.9: Block diagram characterising influenced chaotic encryption. Block 1 represents the chaotic dynamical system where the plaintext $\{p_n : n = 0, 1, \cdots\}$ is an external input. Blocks 2, 3, 4 and 5 perform the same functions as for isolated chaotic encryption and are taken to be identical to blocks 2, 3, 4 and 5 characterised in table 7.1.

second simulation and shows that every subsequence of 8 bits (starting from different initial conditions) occurs with approximately equal probability (and as a consequence all subsequences of less than 8 bits must also occur with equal probability). These results show that the chaotic pseudo-random number generator performs consistently with an ideal pseudo-random number generator, but cannot of course demonstrate complete equivalence because of the necessarily finite lengths of the sequences.

7.4 Influenced Chaotic Encryption

In influenced chaotic encryption each bit of the plaintext perturbs each iterate of a chaotic dynamical system. The output from the dynamical system, after undergoing the same transformations as for isolated chaotic encryption, becomes the cryptogram. The receiver decrypts the cryptogram by performing the inverse transformation to the transmitter. Influenced chaotic encryption is characterised by the block diagram shown in figure 7.9. Block 1 represents the chaotic dynamical system, where the plaintext is an external input. Blocks 2, 3, 4 and 5 represent the same transformations as in isolated chaotic encryption (these transformations are characterised by the functions specified in table 7.1). The output of block 5 in figure 7.9 is the cryptogram (remember, from §7.3, that the output from block 5 in isolated chaotic encryption forms the key stream used to encrypt the plaintext).

The chaotic dynamical system represented by block 1 in figure 7.9 is a recursive equation. The plaintext (assumed to be expressed in a binary alphabet) perturbs each iterate of the chaotic dynamical system, i.e.

$$\begin{aligned} \hat{x}_n &= x_n + g(p_n) \\ x_{n+1} &= f(\hat{x}_n) \end{aligned} \tag{7.17}$$

where $\hat{x}_n$ is the perturbed iterate, p_n represents the n^{th} bit in the plaintext, $g(\cdot)$ is a function characterising the perturbation and $f(\cdot)$ is the nonlinearity characterising the chaotic equation. The perturbing function $g(\cdot)$ studied in this section is given by

$$g(p_n) = \begin{cases} (x_n - 10^{-\alpha}) \bmod(1) & \text{if } p_n = 0 \\ (x_n + 10^{-\alpha}) \bmod(1) & \text{if } p_n = 1 \end{cases} \tag{7.18}$$

where α is an integer (in this section $\alpha = 4$).

The two chaotic recursive equations, (7.4) and (7.5), analysed in the previous section are analysed here as possible candidates for influenced chaotic encryption (i.e. the canonical equation with $g = 4.0$ and the hierarchy of two canonical equations with $g_1 = 3.9$ and $g_2 = 4.0$). Perfect secrecy is obtained when the cryptogram $S_c = \{c_n : n = 0, \cdots\}$ is statistically independent of the plaintext. This is equivalent to requiring the cryptogram to possess maximum entropy for all possible plaintexts. To determine the entropy of the cryptograms generated by the encryption system depicted in figure 7.9, the same two tests described in the final three paragraphs in §7.3 are performed. The first test calculates the relative frequency with which consecutive subsequences of length n occur within the cryptogram. This test consists of the following sequence of computational simulation steps:

1. Choose an initial condition pseudo-randomly (using whatever random function routine is standard for the available computer - VAX Fortran for the simulations reported herein).

2. Use the initial condition chosen in step 1 as a seed for the influenced chaotic encryption scheme. Iterate the encryption scheme 10,000 times (to reduce any initial transients) to generate the cryptogram $\{c_i : i = 1, \cdots, 10,000\}$.

3. Generate a plaintext sequence $\{p_i : i = 1, \cdots, nm\}$ by concatenating m subsequences $S_p(j)$ $j = 1, \cdots, m$ of length n (i.e. $\{p_{(j-1)n+i} : i = 1, \cdots, n\} = S_p(j)$ for $j = 1, \cdots, m$), where each possible subsequence of n bits has a particular (which is detailed below) probability of occurrence.

4. Interpret each $S_p(j)$ as a binary number, where $\{p_i : i = 1\}$ is interpreted as the least significant bit and $\{p_i : i = n\}$ is interpreted as the most significant bit.

5. Plot the interpreted numbers on a histogram, with the vertical axis showing the frequency of occurrence and the horizontal axis showing the ordered numbers $0, \cdots, 2^n - 1$.

6. Generate the cryptogram $\{c_i : i = 10,001, \cdots, nm\}$.

7. Split $\{c_i : i = 10,001, \cdots, nm\}$ into m consecutive subsequences of length n, i.e. $S_c(j) = \{c_{(j-1)n+i} : i = 10,001, \cdots, 10,000 + n\}$ for $j = 1, \cdots, m$.

8. Interpret each $S_c(j)$ as a binary number, where $\{c_i : i = 1\}$ is interpreted as the least significant bit and $\{c_i : i = n\}$ is interpreted as the most significant bit.

9. Plot the interpreted numbers on a histogram, with the vertical axis showing the frequency of occurrence and the horizontal axis showing the ordered numbers $0, \cdots, 2^n - 1$.

The second test calculates the relative frequency with which cryptograms of length n (where each cryptogram of length n is started from a different randomly chosen initial condition) occur. The test consists of the following sequence of computational simulation steps:

1. Choose m initial conditions pseudo-randomly (as in step 1 of the simulation previously discussed in this section).

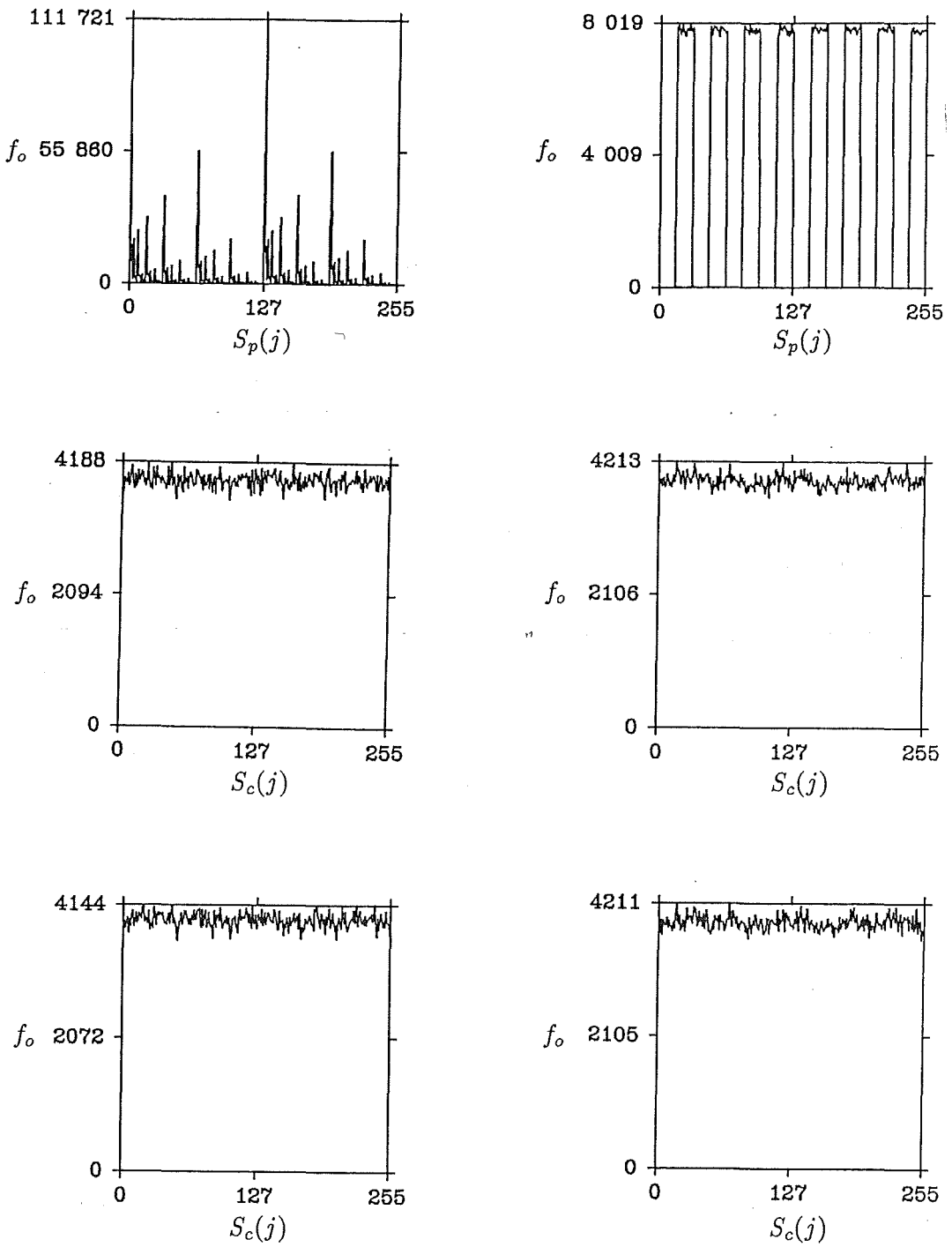


Figure 7.10: Histograms recording the results obtained from the computer simulations for the canonical equation, for $m = 10^6$, $n = 8$ and for two different plaintexts. The horizontal axis of each histogram lists the 256 combinations possible with 8 bits. The vertical axis shows the frequency of occurrence f_o with which each 8 bit sequence occurs. The top row of histograms shows the distribution of the m subsequences $S_p(j)$ for the two plaintexts invoked in the simulations. The middle and bottom row of histograms show the results from the first and second simulation respectively. The left and right hand columns show the results for the plaintext distributions identified at the top of each column. Identical results were obtained for the hierarchy characterised by (7.11) and consequently results for this hierarchy are not shown.

2. Use each of the m initial conditions chosen in step 1 as the seeds for m influenced chaotic encryption schemes. Iterate each of the m encryption schemes 10,000 times (to reduce any initial transients) to generate m cryptograms $\{c_i : i = 1, \cdots, 10,000\}$.

3. Generate a plaintext sequence as outlined in steps 3 to 5 of the simulation previously discussed in this section.

4. Generate the next n bits for each of the m cryptograms, i.e. $S_c(j) = \{c_i : i = 10,001, \cdots, 1,000 + n\}$ for $j = 1, \cdots, m$.

5. Interpret each $S_c(j)$ as a binary number, where $\{c_i : i = 1\}$ is interpreted as the least significant bit and $\{c_i : i = n\}$ is interpreted as the most significant bit.

6. Plot the interpreted numbers on a histogram, with the vertical axis showing the frequency of occurrence and the horizontal axis showing the ordered numbers $0, \cdots, 2^n - 1$.

Histograms recording the results obtained from the computer simulations for the canonical equation, for $m = 10^6$, $n = 8$ and for two different plaintexts are presented in figure 7.10. The top row of histograms shows the distribution of the m subsequences $S_p(j)$ for the two plaintexts invoked in the simulations. The middle and bottom rows of histograms show the results from the first and second simulation respectively. The left and right hand columns show the results for the plaintext distributions identified at the top of each column. Although the plaintext distributions are significantly nonuniform, the distributions plotted in the middle and bottom histograms are approximately uniform, which is consistent with the cryptogram possessing maximum entropy. Additional simulations confirm that as m is increased the amplitude of the wiggles on these distributions decrease, suggesting that the wiggles are the result of statistical sampling.

Further simulations using a wide variety of plaintext statistics have been performed. Plaintexts consisting of repeating subsequences of different lengths (repetition lengths from 1 to 16 bits), and plaintexts consisting of finite length subsequences (lengths from 1 to 16 bits) having nonequal probability of occurrence, were invoked for these simulations. In each case the results were consistent with, but did not confirm that the statistics of the cryptogram are independent of, the plaintext. To prove this conclusively would require the tests to be completed on indefinitely long sequences, which is obviously impractical. Identical results were obtained for the hierarchy characterised by (7.11), and consequently results for this hierarchy are not shown.

7.5 Hardware Implementation Considerations

This section outlines some of the hardware considerations for implementing the chaotic pseudo-random number generator used in isolated chaotic encryption. The modifications required to implement influenced chaotic encryption are not discussed, although they are relatively straight forward to incorporate. Recall that block 1 in figure 7.5 represents a discrete chaotic dynamical system, blocks 2, 3 and 4 in

Figure 7.11: Schematic diagram of a modulo one adder (left hand circuit) and a cyclic modulo one adder (right hand circuit). The square represents an adder, the rectangles represent registers (i.e. storage elements). Addends enter the adder from the right and the sum exists from the left.

figure 7.5 represent cyclic modulo one adders, and block 5 represents a thresholding operation. The hardware implementation of each of these operations is discussed in this section.

An adder is a standard digital hardware component. It inputs two binary numbers (termed addends) and outputs one binary number (termed sum) (*cf.* Taub 1985, Chapter 5). An adder performs arithmetic addition on the two binary addends to form a binary sum. Depending on the design and technology used to construct an adder, addition times of less than 10ns (100 million additions/second) are achievable. A modulo one adder can be formed from any adder if the addends are interpreted as belonging to the interval $[0, 1]$ and if the carry or arithmetic overflow out of the adder is ignored. Figure 7.11 shows the schematic diagram of a modulo one adder and a cyclic modulo one adder (i.e. the sum becomes one of the addends in the next addition). The adder is represented as a square where the addends enter from the left and the sum exits from the right. Memory elements termed registers (represented by rectangles) may be required for the holding (i.e. the storage) of addends during their addition. However, it should be possible through careful design to eliminate the need for some of these registers. The cyclic modulo one adder is formed by feeding the sum output back to one of the addend inputs via a register. The signal termed clock controls the timing of the entire circuit. Modulo one adders are fast, straight forward and inexpensive to implement.

The thresholding operation represented by block 5 in figure 7.5 can be implemented easily. Each binary number outputed from block 4 can be represented as

$$b_n^* = a_{1,n}2^{-1} + a_{2,n}2^{-2} + \cdots + a_{M,n}2^{-2} \tag{7.19}$$

where M is the number of bits in the binary number b_n^* (i.e. the precision of b_n^*) and $a_{i,n}$ is the value of the i^{th} bit in the number b_n^* (i.e. $a_{i,n} = 0$ or 1). The binary output b_n (i.e. the output from block 5) is obtain by selecting the appropriate $a_{i,n}$. The thresholding operation requires no hardware except a connection from the output (of the pseudo-random number generator) to the appropriate bit from the sum output of the cyclic modulo one adder represented by block 4.

The implementation of the chaotic dynamical system is in general more involved. The canonical equation (7.10) requires a multiply, which like addition, is a standard digital hardware component (*cf.* Taub 1985, §5.15). Since the numbers generated by the canonical equation are restricted to the interval $[0, 1]$ the multiply hardware can be simplified, enabling the canonical equation to be implemented with relative ease. However, the inverse canonical equation (7.11) requires a square root, and requires considerably more hardware than a multiply to implement. One possibility

is to implement equation (7.10) and (7.11) in software. However, this would reduce the maximum bit rate obtainable from the pseudo-random number generator to about $10^4 - 10^5$ bits/sec, which is perhaps three orders of magnitude less than if the equations could be implemented directly in hardware. Note that a bit rate of $10^4 - 10^5$ bits/sec is sufficient for many applications (e.g. electronic funds transfer, terminal-computer connections, data acquisition). A second possibility is to implement equation (7.10) and (7.11) directly in hardware using custom made very large scale integrated (VLSI) circuitry. This would enable maximum bit rates of about $10^7 - 10^8$ bits/sec to be obtained, while at the same time maintaining a physically compact construction.

The chaotic dynamical system characterised by (7.10) and (7.11) requires extensive hardware for their implementation. However, other chaotic dynamical systems which are easier to implement exist. For example, the canonical equation for $g = 4.0$ can be transformed into a simpler equation which is easier to implement in hardware. The dynamical properties of the transformed equation however, may not necessarily be the same as those presented in §7.2. If the transformation

$$\bar{x} = \frac{2}{\pi} \sin^{-1} \sqrt{x} \tag{7.20}$$

is applied to (7.10) when $g = 4$, then what is termed the tent map is obtained

$$\bar{f}(\bar{x}) = \begin{cases} 2\bar{x} & \text{if } \bar{x} < 0.5 \\ 2(1 - \bar{x}) & \text{if } \bar{x} \geq 0.5 \end{cases} \tag{7.21}$$

The tent map is simpler to implement in hardware than the canonical equation. A multiply by two on binary numbers is equivalent to shifting the number one bit to the left. The complement operation $(1 - \bar{x})$ is achieved by an operation called twos complement (*cf.* Taub 1985, §8.2), which is a relatively simple logic operation. The inverse of the tent map is

$$\bar{f}^{-1}(\bar{x}) = \begin{cases} \frac{1}{2}\bar{x} & \text{if } \bar{x} < 0.5 \\ 1 - \frac{\bar{x}}{2} & \text{if } \bar{x} \geq 0.5 \end{cases} \tag{7.22}$$

Dividing a binary number by two is equivalent to shifting the number one bit to the right. The complement operation $(1 - \bar{x}/2)$ is obtained by shifting one place to the right and performing a twos complement operation. The operations of shifting and twos complement can be performed using special registers and operation times of less than 10ns are achievable (Taub 1985, §8.2). Using 64 bit binary numbers (i.e. 19 decimal digits), repetition periods of 10^{12} may be obtainable (at a bit rate of 10^6 bits/sec such a sequence would repeat after 11.5 days). In principle the repetition period can be made arbitrary long by using sufficiently high precision binary numbers in the implementation of the encryption scheme.

Chapter 8

Packet Switching

Packet switching is a 'quantised' communication technique. Messages are transmitted into the packet switching network by a source user, and routed through the packet switching network to a destination user. There is a limit to the length of a message a source user can continually transmit into a packet switching network. If a source user wishes to transmit a message longer than the allowed limit, the message must first be quantised into smaller units of information, called *packets* (Kleinrock 1976; Schwartz 1977; Bertsekas and Gallager 1987). The information contained in packets is represented by binary digits (bits). The length of each packet can be less than or equal to the allowed limit. The advantage of doing this is that it enables the network to be more flexible (i.e. enables the network to perform functions impossible in other networks) and more efficient (i.e. utilisation of switching and communication resources is higher by comparison with other networks) (Bertsekas and Gallager 1987).

Figure 8.1 is a representation of a packet switching network. It contains a number of switching nodes (represented by circles) interconnected by communication circuits. A user (represented by squares) connects into the network at a switching node. The network is conventionally depicted within a 'cloud' (i.e. the deformed simple closed curve shown in figure 8.1) because it is rather pointless to attempt to prescribe the exact boundary of the network (the boundary is in fact somewhat arbitrary) (Bertsekas and Gallager 1987, page 3). Packets accepted from a source user are

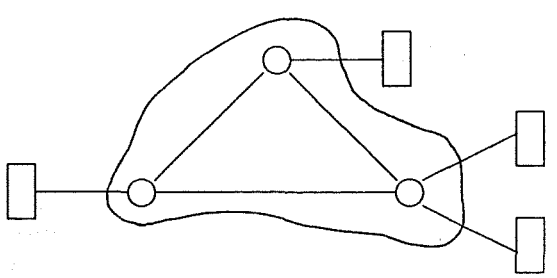


Figure 8.1: A packet switching network containing three switching nodes. The network is contained within a 'cloud', which represents the boundary between the network and the users. Users are external to the network and are connected into the network at a switching node.

routed by the network to a destination user. The destination user reassembles the received packets into the original message (Bertsekas and Gallager 1987). A stream of packets flowing through a communication circuit is termed packet traffic or *traffic.* It is sometimes convenient to refer to the source user and the destination user as a single entity, termed the *source-destination pair.* All items of hardware required for switching and transmitting packets through the network are referred to as *network resources.* The amount of resources in the network is determined by the amount of hardware present and the speed at which it operates.

Packet throughput is a measure of the packet flow rate through a circuit or a network. It is measured by counting the number of packets that pass a packet counter in a certain time interval. If the measuring time interval is relatively short compared to the time it takes a packet to traverse the circuit or network, the measured throughput is termed the *instantaneous packet throughput.* If the measuring time interval is relatively long compared to the time needed for a packet to traverse the circuit or network, the measured throughput is termed the *average packet throughput* or packet throughput, and is denoted tr. The maximum possible packet throughput of a network is termed the *network capacity,* and is determined by the amount of resources within the network. The sojourn time of a packet within a network is defined to be from when the last bit of a packet enters the network until the first bit of a packet exits the network. The average sojourn time of packets within the network is termed the *packet delay.*

The performance of a packet switching network is often specified in terms of the packet delay and the packet throughput (Kleinrock 1976; Reiser 1982). Low packet delay and high packet throughput are descriptors of good performance. Descriptors characterising the performance of a packet switching network are called the grade of service (GOS). The network is said to be congested when the network performance degrades (i.e. when the packet delay increases above some limit and the packet throughput decreases below some limit) (Gerla and Kleinrock 1980). Congestion can occur when the demand (i.e. the number of users and/or the rate at which users send packets into the network) exceeds the network capacity. To prevent congestion some form of packet entry flow control is required (Gerla and Kleinrock 1980). Algorithms invoked to control packet flow are termed flow control algorithms (Gerla and Kleinrock 1980). Most such algorithms require each packet sent into the network to be acknowledged when the packet has been successfully received by the network. If a user does not receive an acknowledgement, the user is usually allowed to send 1 or 2 more packets into the network, but after this the user must wait for an acknowledgement before further packets can be sent. Most flow control algorithms are designed to control the rate at which acknowledgements are sent back to users (Gerla and Kleinrock 1980).

The requirement for ever increasing bandwidth in communication networks runs unabated. The existing design approach is to pack ever more resources into the network, taxing existing technology to the limit. However, an alternative and/or complementary approach is to make use of the network resources more efficiently. One way of achieving this is to encourage users to operate a deterministic packet departure process. This enables the network to predict the future arrival time of packets, allowing resources to be allocated and released more efficiently. The flow control algorithm which is developed in §8.2 encourages users to operate a deterministic packet departure process. However, under certain packet arrival patterns the

flow control algorithm behaves chaotically. This algorithm is based on a modification to an algorithm which forms the basis of many flow control algorithms used in practice (Gerla and Kleinrock 1980). The necessary packet switching concepts and terminology required to describe this algorithm are presented in §8.1.

Managing a packet switching network requires knowledge of the user demand and network capacity. A network is usually considered to be well managed if the network is set close to the user demand (Schwartz 1977). To determine how close a network is operating near its capacity, the network performance should continually be measured. To assess the user demand, and predict future increases, measurement of user traffic statistics are required. The development of a number of instruments suitable for traffic and performance measurement on PACNET, the public packet switching network operated by Telecom Corporation of New Zealand Limited, is presented in §8.3, which also reports results of measurements made with such instruments.

New results are presented in sections §8.2 and §8.3. The conclusions that can be drawn from the material presented in this Chapter, together with suggestions for further work, are included in Chapter 9.

8.1 Overview of Packet Switching

During the early 1950s remote terminals and other peripheral devices began to be connected to centralised computer centres through non-switched communication circuits. From these embryonic stages modern data communication networks have evolved. ARPANET (Kleinrock 1976) and TYMNET (Tymes 1981), introduced around 1970, were the first modern large scale general purpose data networks connecting geographically distributed computer systems. These networks contain switching nodes, various pairs of which are connected by communication circuits. External to each such network are users (e.g. computers, data bases, terminals). They originate messages which pass into the network. These messages pass from node to node on communication circuits, and finally pass out of the network to a destination user. The switching node, to which the source user of a source-destination pair is connected, is termed the *entry node.* The switching node to which the destination user of a source-destination pair is connected, is termed the *exit node.*

Before a packet switching network can accept packets from a source user, an exchange of information must take place between the source user and the network (Bertsekas and Gallager 1987). The purpose of this information exchange is for the source user to inform the network what service is being requested, and for the network to inform the source user if it accepts or rejects the request. In general, the service being requested is the establishment of a connection to a destination user. Such a request is termed a *call request.* The network may reject the call request if it has insufficient spare capacity, or if the destination user is already engaged. If the network accepts the call request, a call is set up, and this is termed a *call.* When the source user and/or the destination user considers the call to have finished, the network is again informed. There are three broad classes of data communication networks (Bertsekas and Gallager 1987):

Circuit switching networks provide communication between pairs of users by establishing a fixed amount of communication bandwidth between them for the

duration of the the call. The connection is set up by a special signalling message which threads its way through the network from the source of the call to the destination.

Message switching networks provide communication between pairs of users by switching and transmitting messages between different switching nodes until the destination user is reached. Messages can be of any length. Individual network nodes receive and store each entire message before passing the message onto the next node in the path to the destination user.

Packet switching networks provide communication between pairs of users by switching and transmitting small units of information called packets between different switching nodes until the destination user is reached. If a source user wishes to transmit a message longer than the allowed packet limit, the message must first be quantised into the smaller units. Individual network nodes receive and store each entire packet before passing the packet onto the next node in the path to the destination user. Packet switching networks have sufficiently important advantages over message switching networks that they are given a different name. These advantages occur because packets are of a finite maximum length as opposed to messages which can be of any length. The two most important advantages of this are: both the memory requirement at each node, and the time required for a message to traverse the network, are less. The reason why a message takes less time to traverse the network is because the entire message (since it is split into packets) does not have to be stored at each node. Thus nodes do not wait for the entire message to be received, but pass on packets (that make up the message) to the next node as soon as they are received.

Since messages and packets are stored in their entirety, before the message or packet is forwarded to the next node in the route to the destination, message and packet switching networks are also termed *store and forward networks* (Bertsekas and Gallager 1987). Message switching networks were developed originally for telegraph switching, and the switching functions within the nodes were performed by people. It was simpler if entire messages were received at each node, before the message was forwarded to the next node in the route to the destination. During the 1960s electronic switching nodes began to be developed. Due to the expense of electronic memory and large message delays (because of the need to store entire messages at nodes before forwarding), message switching evolved into packet switching. The two most important characteristics of store and forward networks are:

- Flexibility in setting up user connections (e.g. the communication circuit connecting the source user and destination user to the network need not have the same transmission speed).
- Efficient use of network resources (e.g. the network resources are shared among all users).

There are essentially two different ways to route packets through a packet switching network (Gerla and Kleinrock 1980). One way is for a route through the network to be established when the source user makes a call request. All packets sent by

the source user (or destination user) follow this route, which remains fixed for the duration of the call. Such a route is called a *virtual circuit.* This method of routing is termed *virtual circuit routing.* The other way, which may or may not involve a call request, is for packets sent into the network to find there own path to the destination user. Different packets from the same source user may follow different routes, and may arrive at the destination in a different order from that in which they set out from the source user. Such packets are termed *datagrams.* This method of routing is termed *datagram routing.*

A path that allows transmission of information in one direction only is termed a *channel.* A path that allows transmission of information in both directions simultaneously is termed a *circuit.* A virtual circuit is composed of a series of channels, termed *logical channels.* Nodes communicate with neighbouring nodes over a logical channel. Each logical channel is associated with one virtual circuit. Many logical channels can traverse a single physical communication circuit. Each logical channel is labelled with a particular integer, termed the *logical channel number.* If two virtual circuits traverse the same physical communication circuit, each virtual circuit is assigned a different logical channel over that communication circuit. Each packet within a network operating virtual circuits, carries (i.e. is labelled with) a logical channel number. At each node, the logical channel number within each packet is updated, to correspond to the next logical channel the packet is about to pass through on the route to the destination. The network requires a certain amount of information to be associated with each packet (e.g. the logical channel number). This information is concatenated onto the front (or head) of the packet, and is termed the *packet header.*

Packet switching networks that operate virtual circuits are theoretically less efficient than networks which operate datagrams. This is because the virtual circuit route remains unchanged for the duration of the call, and therefore cannot take advantage of changing traffic patterns that might occur during the duration of the call. However, virtual circuit networks have a number of practical advantages (Gerla and Kleinrock 1980; Schwartz and Stern 1980):

- Packets arrive at the destination in the same order as sent, so that no resequencing of packets is required at the destination.
- The packet header is smaller since it carries only the logical channel number, whereas datagrams must carry the entire destination address.
- Flow control of individual users is simplified since packets from individual users can be identified.

The purpose of a routing algorithm is to find a route through the network, from the source user to the destination user. Shortest path routing algorithms find routes which minimise some cost. Routing algorithms used in packet switching networks are based on shortest path algorithms, the cost criterion differing among networks (Schwartz and Stern 1980). The most common objective of shortest path routing algorithms is to choose a route which achieves the lowest average network packet delay, and the highest packet throughput. There are two shortest path algorithms which are commonly used in packet switching networks. One, due to Dijkstra (Schwartz and Stern 1980), requires centralized computation. The other, by Ford and Fulkerson

(Gallager 1977; Schwartz and Stern 1980) is suitable for distributed computation. Schwartz (1977) reviews current routing algorithms for packet switching networks (*cf.* Rudin and Mueller 1980; Sproule and Mellor 1981; Tymes 1981). Variable network conditions, such as line failures and changing traffic patterns, make it necessary for routing algorithms to be adaptive in order to best achieve their objectives (Muralidhar 1984).

Network congestion can occur when the user demand exceeds the network capacity. Network congestion leads to large packet delays and reduced packet throughput. Since a packet switching network shares resources among many users, there is always a possibility of network congestion (Sproule and Mellor 1981). It is important to note that, although routing algorithms can reduce and perhaps delay network congestion, they cannot prevent congestion (Schwartz and Stern 1980). The control of packet flow into the network is essential for preventing network congestion (Gerla and Kleinrock 1980). Algorithms which control the flow of packets into the network, and within the network, are termed *flow control algorithms.* Interactions between routing and flow control algorithms have also become an active research topic. The main functions of a flow control algorithm are to (Gerla and Kleinrock 1980):

- Prevent degradation of packet delay and throughput.
- Allocate resources equitably among users.
- Ensure that the rate at which packets enter the network from the source is the same as the rate at which packets exit the network to the destination. This is termed speed matching between source-destination pairs.

There are three different flow control levels within any network, although a single algorithm may operate at all levels (Georganas 1980; Gerla and Kleinrock 1980). These levels are described in the next four subsections. Kleinrock (1976), Lam (1976) and Jaffe (1981) describe static flow control algorithms. Kermani and Kleinrock (1980) and Jain (1986) discuss dynamic flow control algorithms, while Rudin and Mueller (1980) and Muralidhar (1986) investigate interactions between routing and flow control.

8.1.1 Hop Level Flow Control

The technique known as hop level flow control has the objective of controlling the flow of packets entering switching nodes. On entering a node, packets are usually separated into different groups or classes. Hop level flow control assign node resources (e.g. memory, processing time) between the different classes of packets. A number of different hop level flow control algorithms are used in practice, the most common being:

Channel queue limit algorithm. Incoming packets to a node are separated into classes according to the output queue or channel the packets are destined for. Flow control is achieved by ensuring that no output queue exceeds a certain limit. Different classes may have different output queue buffer limits. For example ARPANET (Kleinrock 1976) uses a channel queue limit algorithm. Output memory buffers are shared among the classes, with each class having a minimum and a maximum allocation limit.

Buffer class algorithm. Incoming packets are separated into classes according to the packet hop count (i.e. the number of communication circuits that the packets have traversed so far). This means each node keeps track of $n-1$ (where n is the number of the nodes in the network) classes of traffic and allocates a fixed or dynamically changing number of buffers to each packet class. GMDNET (Gerla and Kleinrock 1980) implements a buffer class flow control algorithm.

Virtual circuit algorithm. Incoming packets are separated into classes according to the virtual circuit on which the packet is received. This algorithm can only be implemented on networks that operate virtual circuits. The number of classes varies with time, since the number of virtual circuits in the network varies with time. TRANSPAC (Gerla and Kleinrock 1980), TYMNET (Tymes 1981) and PACNET are examples that use a virtual circuit flow control algorithm.

Apart from the hop level flow control algorithms outlined above, there are many other possible hop level flow control techniques. For example, packet classes may be distinguished on the basis of packet source, destination or source-destination pair. However, the three algorithms listed above are the only algorithms which have been analysed extensively and implemented in real networks.

8.1.2 Network Access Flow Control

The technique known as network access flow control has been devised to control the flow of packets entering a network from users. The state of internal network congestion determines the degree to which the flow of packets is throttled back. Measurements of congestion may be local (e.g. based on buffer occupancy), global (e.g. based on the total number of packets inside the network), or selective (e.g. based on source-destination pairs). A number of network access flow control algorithms are used in practice, the most common ones are:

Isarithmic algorithm. This is based on the concept of a permit (i.e. a ticket that permits a packet to travel from a source to a destination). Under this concept the network is initially provided with a number of permits, several are held in store at each node. Each packet accepted into the network acquires one permit. This results in a reduction of permits available at the entry node. The accepted packet traverses the network until its exit node is reached. The exit node receives the permit and adds it to its permit storage. To avoid permits accumulating at any node, precautions are taken to limit the number of permits at any node. If a newly released permit cannot be accommodated in a node it is forwarded to another node where it could be accommodated.

Input buffer limit algorithm. This distinguishes between user input packets and transit packets. It throttles the input traffic based on buffer occupancy at the entry node. It favours transit packets over user input packets, which is a desirable property since the network has already invested resources in transit packets. Many different input buffer algorithms may be contemplated. For example, the number of input packets may be limited to a maximum.

Choke packet algorithm. This is based on the notion of link (connections between nodes) and path (a route between a source and a destination) congestion. A

link is said to be congested if its utilization exceeds a given limit. A path is congested if any links along the path are congested. When a node receives a packet directed to a destination whose path is congested, a special packet called a choke packet is sent back to the entry node. The choke packet informs the entry node that the path to that destination is congested and instructs it to block any subsequent input packets to that destination. The path to the destination is assumed to have become clear when no choke packets have been received for a specific time period.

8.1.3 Entry to Exit Flow Control

The objective of entry-to-exit flow control is to prevent congestion at the exit node of the network. This can occur if the source user transmits packets at a higher rate than can be accepted by the destination user (Kumar 1980). This level of flow control usually uses packet acknowledgements to effect flow control. The exit node transmits an acknowledgement back to the entry node, but does so only after it has transmitted a packet to the destination user. An upper bound is set on the number of packets that can be transmitted by the entry node and which have not yet been acknowledged by the exit node. This upper bound (a positive integer) is called the *window size* (or window). Upon receiving an acknowledgement, the entry node is free to transmit one more packet to the exit node. Acknowledgements are either transmitted as special control packets or are piggybacked on regular packets. While the main object of entry-to-exit flow control is to prevent congestion at the exit node, an important by-product is the prevention of global congestion.

8.2 Flow Control Exhibiting Deterministic Chaos

The requirement for ever increasing bandwidth in communication networks remains unabated. Established design approaches for achieving this are to pack ever more resources into the network, taxing existing technology to the limit. However, an alternative and/or complementary approach is to make use of network resources more efficiently. One way of achieving this is to encourage users to operate a deterministic packet departure process. This enables the network to predict the future arrival time of packets, allowing resources to be allocated and released more efficiently. The flow control algorithm developed in this section encourages users (by offering an improved grade of service (GOS) to the user) to operate a deterministic packet departure process.

Encouraging certain types of user behaviour represents a means of improving the utilisation of resources within a communication network. However, it also results in the user behaviour being influenced or controlled by the network. Flow control provides a feedback mechanism by which chaotic or other forms of self-organising behaviour may occur (refer to §3.2). Certain forms of such behaviour may maximise the utilisation of network resources. The development of a packet entry flow control algorithm that induces resource efficient self-organising behaviour, while at the same time delivering the specified GOS to users, may inspire considerable advances in communication network design. The flow control algorithm introduced in this section, termed 'cooperative flow control', is my attempt at achieving this.

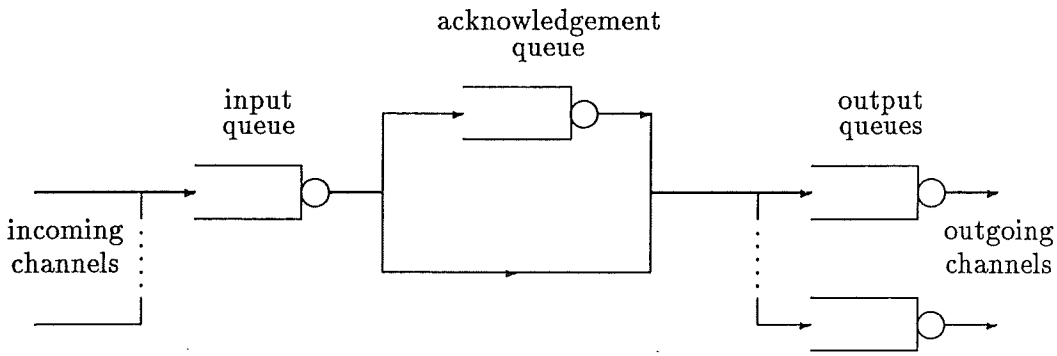


Figure 8.2: The switching node queue structure required for the implementation of cooperative flow control. There are n incoming and n outgoing communication channels to the switching node. Each user is connected to the switching node via at least one circuit.

Cooperative flow control is based on a modification to an algorithm which forms the basis of many flow control algorithms used in practice (Gerla and Kleinrock 1980). The algorithm is an input buffer limit flow control algorithm which is suitable for hop level and/or network access level within a virtual circuit packet switching network. The algorithm determines the rate at which acknowledgement packets are sent back to users. The acknowledgement rate is determined from the rate at which packets are received by the network from users. An intriguing and possibly technologically significant feature of this algorithm is that under certain packet arrival patterns the flow control algorithm behaves chaoticly.

The main purpose of a switching node is to provide switching services to entering packets. If there is more than one packet requesting the same service, these packets must queue for the service. The switching node packet queue structure required for the implementation of cooperative flow control is illustrated in figure 8.2. There are n incoming (to the node) and n outgoing (from the node) communication channels. Each user is connected to the network via at least one circuit (i.e. one incoming and one outgoing channel). A queue is represented in figure 8.2 by a rectangle with the left hand edge removed. The right hand edge represents the head of the queue. A server is represented by a small circle at the head of the queue. Arrowed lines indicate the direction of packet flow within the node.

All incoming (user data) packets queue for switching service at the input queue on a first come first serve (FCFS) basis. The switching server processes each packet and determines which outgoing channel each packet should be switched to. Once this decision is made, the packet is transferred to the appropriate output queue. Simultaneously, an acknowledgement packet is generated and placed in the acknowledgement queue on a FCFS basis. The switching server provides the same service to each incoming user packet. The amount of time required to service each packet is constant and is designated μ. Each logical channel has its corresponding acknowledgement queue. The acknowledgement server merely introduces a time delay T_{ack}. After waiting the appropriate time in the acknowledgement queue, the acknowledgement is transferred to the appropriate output queue (i.e. the output queue which transmits back to the user). A buffer limit b_t is applied to the input queue. If the number of packets queuing for service exceeds the buffer limit b_t, no acknowledgements are sent

back to users (i.e. the acknowledgement server is effectively deactivated). When the queue falls below the buffer threshold b_t, the acknowledgement server is reactivated.

Each logical channel is assigned a maximum number of packets which can be sent into the network without any acknowledgements being received from the network. This maximum number is termed the window size w. If w packets have been sent into the network and no acknowledgements have been received, the user is prevented form sending further packets. This is termed *network blocking* or *blocking.*

The time interval between two consecutive packets arriving at a node or a user is termed *interarrival time.* The time interval between two consecutive packets departing from a node or a user is termed *interdeparture time.* Consider a packet stream flowing through a logical channel operating over a communication circuit between a user and a switching node. The n^{th} packet arrives at the node from the user at time A_n, with the interarrival time between the $(n-1)^{th}$ packet and the n^{th} packet being $Ia_n = A_n - A_{n-1}$. The n^{th} packet departs the switching server at time D_n. The interdeparture time between the $(n-1)^{th}$ packet and the n^{th} packet from the switching server is $Id_n = D_n - D_{n-1}$. The n^{th} acknowledgement departs the acknowledgement queue at time $Dack_n$. The interdeparture time between the $(n-1)^{th}$ acknowledgement and the n^{th} acknowledgement from the acknowledgement server is $Idack_n = Dack_n - Dack_{n-1}$. Invoking the sequence notation introduced in §2.1, the sequences of interarrival and interdeparture times are conveniently denoted by

$$\begin{aligned} S_{Ia} &= \{Ia_n : n = 0, 1, \cdots\} \\ S_{Id} &= \{Id_n : n = 0, 1, \cdots\} \\ S_{Iack} &= \{Iack_n : n = 0, 1, \cdots\} \end{aligned} \tag{8.1}$$

The purpose of the acknowledgement server is to introduce the delay T_{ack}, which is a function of the interarrival time between packets flowing over a particular logical channel to the switching node, i.e.

$$T_{ack} = \frac{\alpha}{(Ia_n)^\beta}, \tag{8.2}$$

where α and β are either constant or are made dependent on the network congestion. In what follows α and β are set to the constant values: $\alpha = 1$ and $\beta = 2$.

Cooperative flow control enables the user to make a tradeoff between instantaneous packet throughput and average packet throughput. A user can send one packet immediately after the other for w packets into the network. This achieves the highest instantaneous packet throughput possible. However, because $T_{ack} = 1/Ia_n^2$, then T_{ack} is large (since Ia_n is small). Thus the user is blocked from sending the $(w+1)^{th}$ packet for some time, giving a poor average (long term) throughput. To achieve a high instantaneous throughput the user pays a penalty in average packet throughput. However, if a user sends packets into the network separated by a suitable time interval (time interval = 1 if $\alpha = 1$), the highest average packet throughput is obtained. This allow users to make a definite decision as to what strategy best suits their particular needs. For example, a long file transfer is best achieved by suitably separating the packets in time. An enquiry-response is best achieved by sending packets one immediately after the other.

Penalising users for operating an inappropriate packet arrival strategy promotes good behaviour by users. For a user to achieve the best performance from the network

a particular packet arrival strategy must be used. A user operating an inappropriate packet arrival strategy is (and possibly severely) penalised. Large file transfers for example require the network to allocate resources for a long time. The most efficient way to handle a file transfer is for the network to permanently allocate a small potion of its resources to the transfer, and for those resources to be continually in use. This requires the network and the user to form a partnership, each helping the other. The user must provide a regular supply of packets at an appropriate rate while the network must permanently allocate a portion of its resources. To handle bursty traffic (i.e. traffic which consists of small clusters of packets separated by relatively long periods of no transmission) efficiently the network must provide a pool of resources that can be called upon to handle a burst of packets from any user. Some or all of these resources are idle at times and therefore bursty traffic cannot be handled as efficiently as traffic which arrives at regular intervals. Cooperative flow control discourages bursty traffic and encourages packets to be sent at regular time intervals.

In most packet switching networks, the window size is chosen to ensure queue lengths do not grow too large (and hence introduce large delay), and to ensure acknowledgements get back to users in sufficient time to prevent unnecessary network blocking. A window size of 2 or 3 is usually chosen to fulfill these requirements. In cooperative flow control, queue lengths do not grow significantly as the window size is increased. This is because rapid packet arrival rates are penalised. Thus a larger window size can be used with little increase in packet delay. A larger window size allows bursty traffic to be better accommodated. In §8.3, it is demonstrated that the average packet burst length is about 5 packets for enquiry response traffic. Thus a window size of 5 or greater is preferable. The departure time D_n of packets from the switching server is given by

$$D_n = \sum_{i=1}^{n} Id_i, \quad n = 1, 2, \cdots \tag{8.3}$$

where

$$Id_n = (A_n - D_{n-1})\, U(A_n - D_{n-1}) + \mu, \quad n = 1, 2, \cdots \tag{8.4}$$

where μ is the service time of the switching server, and $U(\cdot)$ is the unit step function

$$U(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases} \tag{8.5}$$

In this section μ is made a constant, $\mu = 1$. The departure time of acknowledgements $Dack_n$ from the switching node is given by

$$Dack_n = \sum_{i=1}^{n} Idack_i, \quad n = 1, 2, \cdots \tag{8.6}$$

where

$$Idack_n = (A_n - Dack_{n-1})\, U(A_n - Dack_{n-1}) + T_{ack}, \quad n = 1, 2, \cdots \tag{8.7}$$

The window ensures that each user cannot send more than w packets into the network without receiving acknowledgements. Therefore, (8.6) is constrained by

$$ql(t) = \int_0^t \sum_{i=1}^{\infty} (\delta(A_i - \tau) + \delta(Dack_i - \tau))\, d\tau \tag{8.8}$$

where $0 \leq ql(t) \leq w$, $ql(t)$ is the length of the queue at time t, and $\delta(\cdot)$ is the dirac delta function (refer to §2.3). Recall that no acknowledgements are sent to users when the number of packets queuing for service exceeds the input buffer limit b_t.

The time duration between the generation of the $(n-1)^{th}$ packet and the n^{th} packet by a user is termed the intergeneration time, and is denoted Ig_n. A sequence of intergeneration times is denoted $S_{Ig} = \{Ig_n : n = 0, 1, \cdots\}$. The interarrival time between packets entering the network is in general different from the intergeneration time, because the network can block the entry of packets. The packets generated by a user while blocked by the network, can be handled in a number of alternative ways. These could include:

- When blocking occurs, delay the generation of future packets by the time during which the user is blocked.
- If a packet is generated while the network is blocked the packet is discarded (but without prejudice to future packet arrivals).
- Requires users to arrange their packets in personal queues before transmitting them. If a packet is generated while the network is blocked, the packet enters the queue and is sent immediately the network becomes unblocked.

Of these possible responses to network blocking, only the first method is examined here. Thus, the interarrival time of packets to the node is related to packet intergeneration time by

$$Ia_n = \begin{cases} Ig_n, \text{ if no network blocking} \\ \text{time at which network blocking ended } - A_{n-1}, \text{ if blocked} \end{cases} \tag{8.9}$$

The packet throughput of a single node is now examined. Assume a user offers the network a stream of packets, where the interarrival time between all the packets in the stream is constant (i.e. $Ia_i = constant$ for $i = 0, 1, \cdots$). If $Ia_i \geq 1/Ia_i^2$ (i.e. $Ia_i \geq 1$), then the network supplies sufficient acknowledgements to the user to prevent blocking, and the entire offered traffic is carried by the network. However, for $Ia_i < 1/Ia_i^2$ (i.e. $Ia_i < 1$), the network blocks the user from sending packets some of the time (i.e. the node cannot carry the offered traffic). Maximum possible packet throughput is obtained when

$$Ia_i = \frac{1}{Ia_i^2} = 1 \tag{8.10}$$

For $Ia_i = constant$ and $Ia_i \geq 1/Ia_i^2$, the packet throughput tp of the logical channel is $tp = 1/Ia_i$. For $Ia_i < 1/Ia_i^2$, tp rapidly falls as $Ia_i \to 0$. The left hand curve in figure 8.3 is a plot of packet throughput against constant packet interarrival time Ia_i (note that the part of the curve for $Ia_i < 1/Ia_i^2$ is obtained by computer simulation).

The middle curve in figure 8.3 is a plot of packet delay through the node against constant packet interarrival time. The right hand curve in figure 8.3 is a plot of packet delay from when the user wants to send a packet until the packet exits the node (i.e. delay through the network added to the packet waiting time at the user) against constant packet interarrival time. A packet is forced to wait in the user's personal queue if the node blocks the user from transmitting. The middle and right

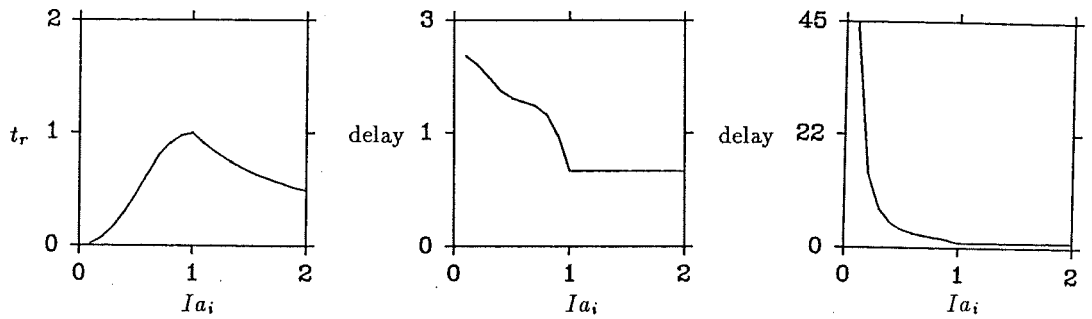


Figure 8.3: Curves characterising packet throughput and delay for a single node. The left hand curve is a plot of packet throughput against constant packet interarrival time. The middle curve is a plot of packet delay through the node against constant packet interarrival time. The right hand curve is a plot of packet delay through the network plus the time packets wait at the user due to network blocking, against constant packet interarrival time. These curves are obtained from a computer simulation.

hand curves in figure 8.3 coincide for $Ia_i \geq 1/Ia_i^2$. For $Ia_i < 1/Ia_i^2$ the packet throughput is poor and packets spend most of their time waiting at the user. Once the node allows a packet to enter the node the packet spends little time traversing the node. The right hand curve in figure 8.3 shows that the delay rapidly increases as $Ia_i \rightarrow 0$, while the middle curve flattens out. If a user wishes to transfer a large file then the best strategy (i.e. to obtain maximum average packet throughput and a low packet delay) is to send packets separated in time such that $Ia_i = 1$.

A packet switching network is a driven deterministic dynamical system. It controls the flow of packets into the network. Flow control provides a feedback mechanism by which chaotic behaviour can arise. Input packets provide the equivalent of the driving force used in driven dissipative dynamical systems. It is usually assumed that the generation of packets by users is a stochastic process. The response of a deterministic dynamical system driven by a stochastic process is in general stochastic (i.e. packets exiting the network appear to form a stochastic process). To analysis the dynamics of a driven dynamical system in terms of power spectra and bifurcation diagrams, it is necessary to drive the network from a source that generates packets in some regular repeating manner. A particularly simple driving waveform is a sinusoid. This section examines the dynamics of a single switching node when Ig_n is given by

$$Ig_n = A \sin\left(2\pi \frac{n}{P}\right) + O \tag{8.11}$$

where Ig_n is the intergeneration time between the $(n-1)^{th}$ packet and the n^{th} packet, A is the amplitude of the sinusoid, P is the period of the sinusoid and O is an offset. The sinusoidal source is specified by the three parameters, A, P and O. Many computer simulations of a single node driven from a sinusoidal packet source has been undertaken. It has been found that the network response is similar for a wide range of A and P. The computer simulations, whose results are presented below, are characterised by:

- One user is connected to the network and operates one logical channel.

- The input buffer threshold b_t exceeds the window size w.
- The window size is set to $w = 9$.
- The sinusoidal packet generation parameters are set to $A = 1$ and $P = 11$.
- When blocking occurs all input packets are delayed until an acknowledgement is received from the network.
- The acknowledgement time period $T_{ack} = 1/I_a n^2$ (i.e. $\alpha = 1$ and $\beta = 2$).

Figure 8.4(a)-(d) show respectively, a typical portion of the sequences S_{Ig}, S_{Ia}, S_{Id}, S_{Iack}, for $O = 1.25$. The value of the n^{th} member of each sequence (i.e. the duration of the n^{th} intergeneration, interarrival or interdeparture time) is represented by the length of the vertical line at the n^{th} position along the horizontal axis. A casual inspection of the sequence S_{Ia} would suggest it is the same as sequence S_{Ig}. However, on closer inspection, the sequence $\{Ia_{iP} : i = 0, 1, \cdots\}$ differs from $\{Ig_{iP} : i = 0, 1, \cdots\}$ (e.g. Ig_{3P} and Ig_{11P} differs from Ia_{3P} and Ia_{11P}; these positions are marked with arrows on figure 8.4), Ia_{3P} and Ia_{11P} are of longer duration than Ig_{3P} and Ig_{11P}. Define the sequence $\{x_n : n = 0, 1, \cdots\}$ to be

$$\begin{aligned} x_n &= Ia_{6+nP} \\ S_x &= \{x_n : n = 0, 1, \cdots\} \end{aligned} \tag{8.12}$$

where P is the period of the packet generation sinusoidal waveform. Figure 8.4(e) shows part of the sequence S_x. Note that, the members of the sequence S_x are not all identical, as is the case for $\{Ig_{iP} : i = 0, 1, \cdots\}$. Figure 8.4(f) shows a typical portion of the length of the effective queue (the number of packets in the input queue plus the number of acknowledgements in the acknowledgement queue) plotted against time. The length of the effective queue is represented by the height of the curve (which is stepped) above the horizontal axis (the maximum length of the effective queue is $w + 1$). Figure 8.4(e)-(f) appear to contain no apparent pattern and do not repeat within the interval shown on the figure.

Figure 8.5 shows two bifurcation diagrams for the sequence S_x (refer to §1.1.6). The top bifurcation diagram is obtained for O over the domain 1.39 to 1.415. The bottom bifurcation diagram gives an expanded view for O over the domain 1.407 to 1.415. Figure 8.6 shows six power spectra of S_x for the O parameter specified in its caption. Arrows indicate the position of these parameter values in relation to the bottom bifurcation diagram shown in figure 8.5. The bifurcation diagram contains extraordinarily rich behaviour. Between regions of apparent chaos are regions of regularity. The power spectra confirm there exist intervals of O for which the sequence S_x has an apparent continuous spectrum separated by intervals of O for which the sequence S_x contains a small number of definite frequencies.

An event is defined to have occurred when a packet arrives or departs the switching node. The time of the n^{th} event is denoted E_n. The time duration between event $n-1$ and event n is denoted $Ie_n = E_n - En - 1$, and the sequence $\{Ie_n : n = 0, 1, \cdots\}$ is denoted S_{Ie}.

A 2-dimensional state space for the switching node can be constructed. A point on a trajectory is plotted in state space at the coordinate (Ie_n, Ie_{n+1}). A trajectory

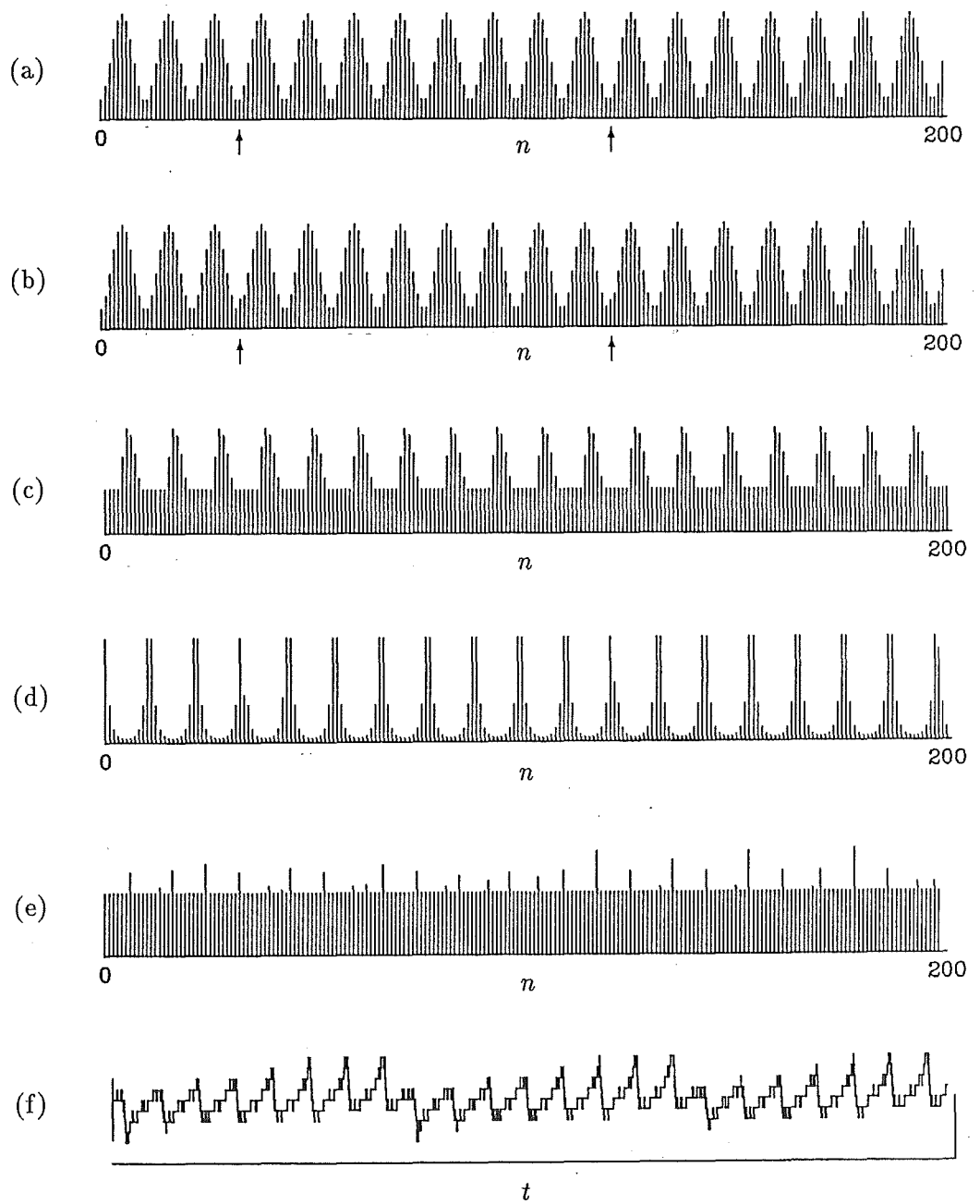


Figure 8.4: Sequences which characterise the switching node. **(a)-(d)** Shows respectively portions of the sequences S_{Ig}, S_{Ia}, S_{Td}, S_{Iack} (for $n = 1, \cdots, 200$) for $O = 1.25$. The duration of the n^{th} intergeneration, interarrival or interdeparture time is represented by the length of the vertical line at the n^{th} position along the horizontal axis. **(e)** Typical portion of the sequence S_x. Note that the members of the sequence S_x are not all equal in value, as is the case for the sequence $\{Ig_{iP} : i = 0, 1, \cdots\}$. **(f)** Typical portion of the effective queue length (i.e. the length of the input queue plus the length of the acknowledgement queue) with time. The horizontal axis forms the time axis. The effective queue length is represented by the height of the stepped curve above the time axis.

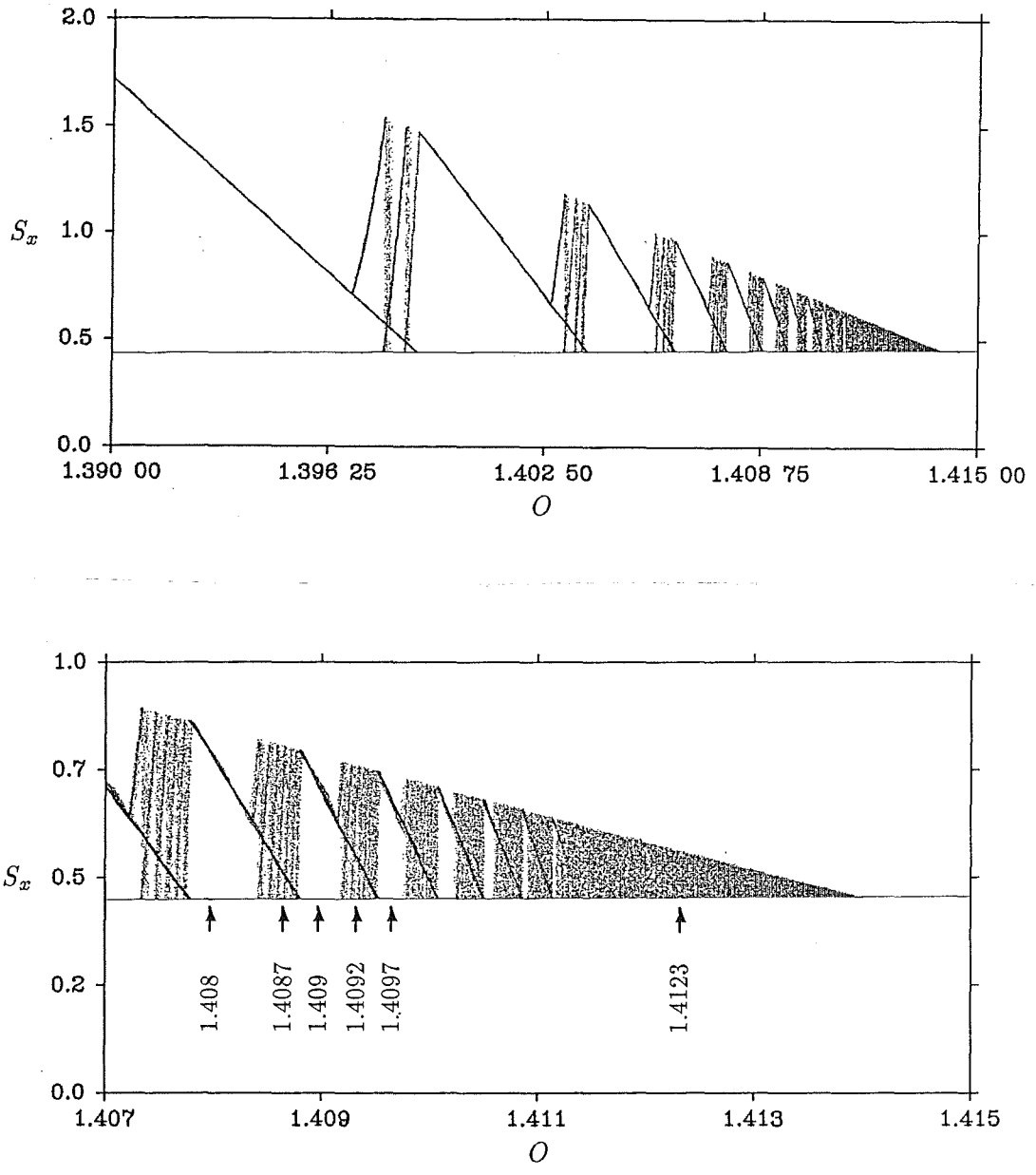


Figure 8.5: Bifurcation diagram of the sequence S_x. The top bifurcation diagram is for O in the domain 1.39 to 1.415. The bottom bifurcation diagram provides an expanded view for O in the domain 1.407 to 1.415.

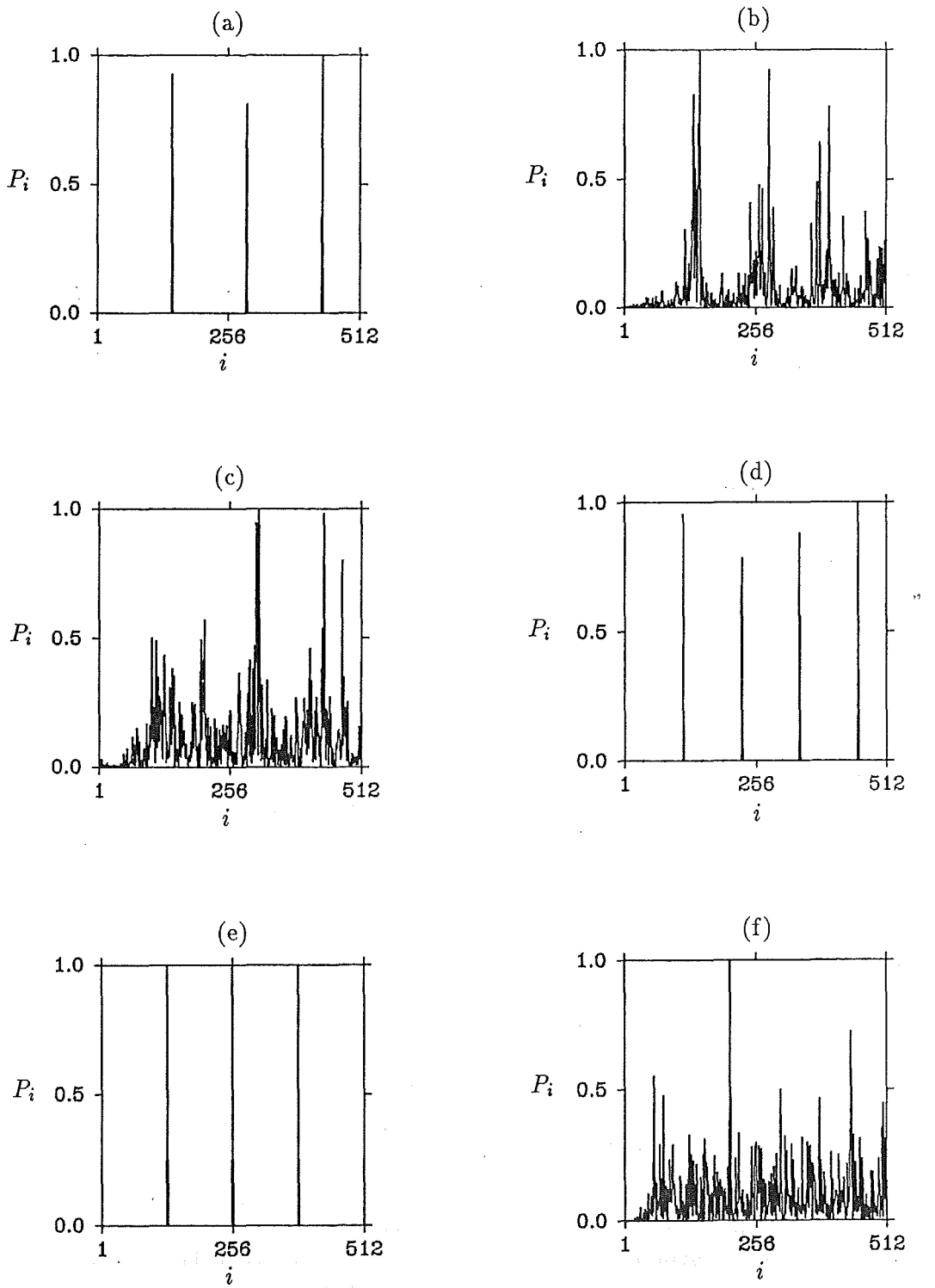


Figure 8.6: Six power spectra of S_x for $A = 1$, $P = 11$ and for the the following values of O: **(a)** $O = 1.407$. **(b)** $O = 1.4087$. **(c)** $O = 1.409$. **(d)** $O = 1.4092$. (e) $O = 1.4097$. **(f)** $O = 1.4123$. Arrows indicate the position of these parameter values in relation to the bottom bifurcation diagram shown in figure 8.5.

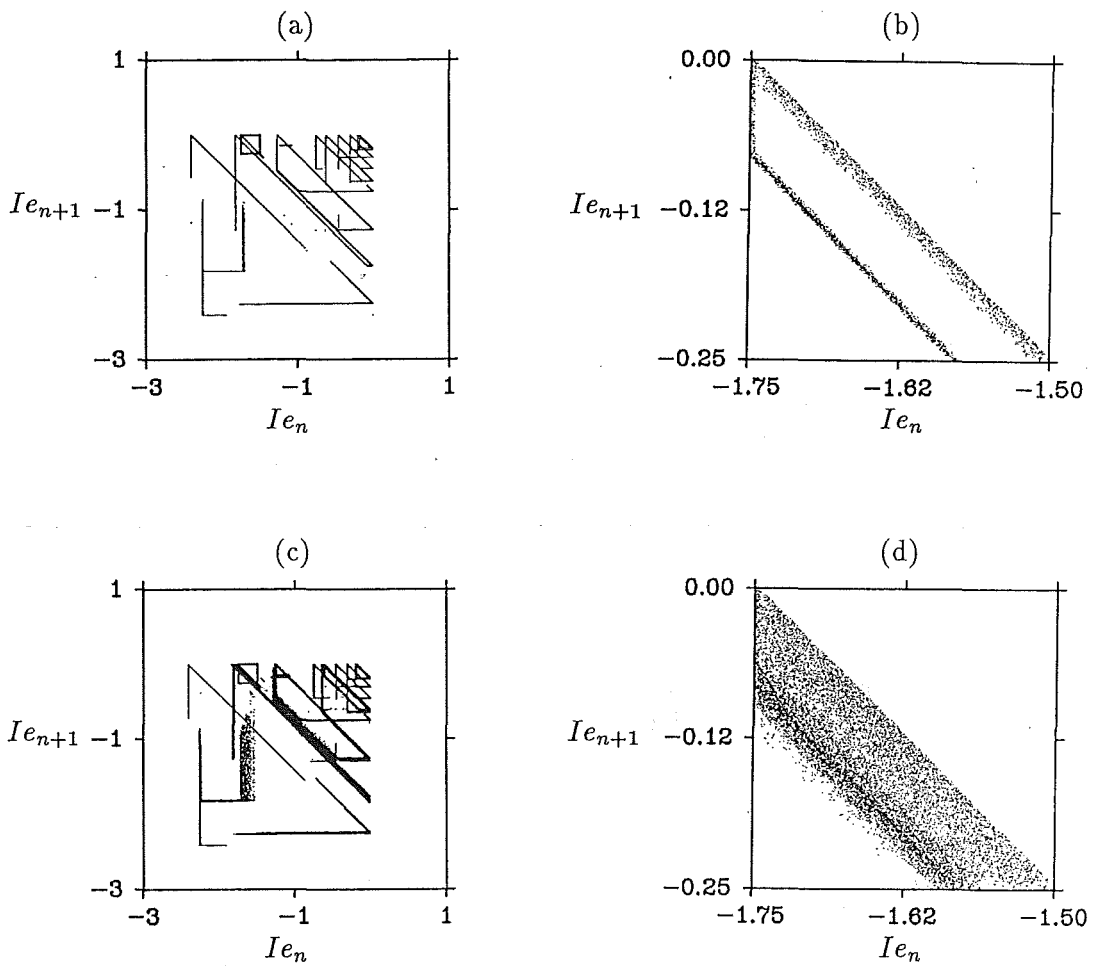


Figure 8.7: Construction of a state space for the switching node. **(a)** Trajectory constructed by plotting 100,000 points. The trajectory is confined to regions in state space that resemble line segments. **(b)** A blowup of a small segment of one of the lines shown in (a). The blowup is obtained by plotting only those points that lie within the small box drawn in the state space of (a). The coordinate of the bottom left hand corner and the top right hand corner of the small box are (-1.75,-0.25) and (-1.5,0.0) respectively. **(c)** Trajectory constructed by plotting 500,000 points. **(d)** A blowup of a small segment of one of the lines shown in (c). The blowup has been obtained by plotting only those points that lie within the small box (in the same position as in (a)) drawn in the state space of (c).

is constructed by plotting the points (Ie_n, Ie_{n+1}) for $n = 0, 1, \cdots$. Figure 8.7(a) shows a trajectory constructed from 100,000 points. The trajectory appears to be confined to regions in state space that resemble line segments. A blowup of a small segment of one of these lines is shown in figure 8.7(b). Figure 8.7(b) is constructed by plotting only those points that lie within the small box (whose bottom left and top right coordinates are (-1.75,-0.25) and (-1.5,0.0) respectively) drawn in figure 8.7(a). Figure 8.7(b) shows that the line segments have a definite width. Figure 8.7(c) shows the trajectory constructed from 500,000 points. A blowup of the same small region is shown in figure 8.7(d). The attractor thickens as the number of points increase, indicating that the trajectory does not repeat. The blowups of figure 8.7(b) and figure 8.7(d) do not reveal self similar scaling, consequently the attractor does not form a fractal. This may indicate the attractor is not confined to a 2-dimensional state space.

To analyse and take advantage of self-organisation in communication networks requires the development of new design and analysis techniques. Self-organised behaviour is holistic, in the sense that it is the whole activity that is of interest, not the state of any specific subsystem (i.e. it is not possible to reduce the system, or a pattern of activity of the system, to the sum of activities of individual subsystems). It may eventually prove feasible to design a communication network to operate in a specific self-organising coherent or chaotic way to enhance network performance and improve network fault tolerance.

8.3 Traffic and Performance Measurement on PACNET

PACNET is a low speed public packet switching network operated by Telecom Corporation of New Zealand Limited (here after called Telecom). Two instruments (developed by the author in collaboration with others) suitable for measuring user traffic statistics and the performance of PACNET is described in §8.3.1 (*cf.* Swan 1985; Silvestor 1985; McCulloch 1986; Murch *et al.* 1987). Before the statistics of a stream of packets can be measured it is necessary to develop a model for the packet generation process (Jain and Pouther 1986). A traffic model characterises traffic statistics with a number of random variables. A suitable traffic model for PACNET users has been developed and is described in §8.3.2. Measuring traffic statistics means recording sufficient information so that the probability distribution functions of the random variables specified in the traffic model can be estimated (Dudick *et al.* 1971). The author has measured traffic on four user connections (i.e. a communication circuit connecting a user into the network) to PACNET. The results of these measurements are presented in §8.3.3. Measurements have been obtained for a Videotex connection (Videotex is a service which enables a user to have access to a number of different computer databases) and two separate business connections to PACNET. The traffic from one business is measured at both a host computer connection and one of its computer terminal connections. The traffic at the other business is measured at only a computer terminal connection to PACNET. These four connections are hereafter referred to as A, B, C and D respectively.

The performance of PACNET is specified by a set of descriptors called the grade of service (GOS). Each descriptor in the GOS is specified probabilistically. For

example, the packet delay through the network, measured from when the last bit of a packet enters the network to when the first bit exits the network, is specified by the mean delay (400ms), and the 95 percentile delay (600ms).

The resources required in a packet switching network depend on the user traffic statistics and on the GOS. In order to ensure that the correct amount of resources are present, it is necessary to continually assess the network performance. Users are split into groups or traffic classes depending on their traffic statistics. To predict the amount of network resources required in the future, it is necessary to know the statistics of each traffic class, and to predict the amount of future increase or decrease in traffic within each traffic class. The prediction of future increases or decreases in traffic within each traffic class is not considered here. Telecoms' ultimate aim is three fold: 1) to accurately measure the performance of PACNET, 2) to determine the statistics of each traffic class, and 3) to investigate how traffic statistics affect the amount of resources required in the network. This section describes and demonstrates instruments capable of achieving aims 1) and 2). These instruments are currently being used by Telecom for performance and traffic measurements on PACNET.

The CCITT (telecommunications consultative committee to the United Nations) has specified a standard protocol to operate between packet switching networks and users. This protocol is called X25 (*cf.* Bertsekas and Gallager 1987, §2.8.2). In X25 terminology the user equipment is termed the data terminal equipment (DTE), and the point where the user connects into the network is termed the data communication equipment (DCE). The X25 protocol defines many different packets, each serving a different function. Only two are of relevance of this section. These are termed the acknowledgement packet and the data packet. An acknowledgement packet is a special control packet that carries an acknowledgement. A data packet carries user data, and performs the same function as the type of packet defined in the first paragraph of this Chapter.

8.3.1 Measuring Instruments

To be able to measure packet throughput and delay, a device termed an echo/absorb board has been constructed (Swan 1985; Silvestor 1985). It can echo (i.e. retransmit each data packet received, back to the source) or absorb (i.e. transmit back an acknowledgement for each data packet received) data packets sent to it. The echo or absorb mode is selected by switches. The echo/absorb board consists of an 8085 microprocessor, an Intel 8274 multiprotocol controller, 8K of RAM and 8K of ROM. Synchronism of data transfer between the 8085 and 8274 utilises interrupts. A maximum line speed of 48,000 bits per second (bps) can be achieved. Although the echo/absorb board only absorbs or echos data packets, it must perform most of the functions required by X25 data terminal equipment (DTE) (i.e. the user). Future developments for the echo/absorb board include: a traffic generation mode, and mode selection carried out on the reception of special packets instead of switches.

For the controlled generation of traffic for delay and throughput measurements on PACNET a VAX 11/750 computer running under Packet Switch Interface (PSI) software is used. Packet delay is measured by probing (i.e. transmitting a packet into the network) the network with a packet, echoing the packet off the echo/absorb board, which has a fixed known delay, back to the source (i.e. the VAX 11/750), and measuring the round trip delay. Packet throughput is determined by measuring how

many packets per second can be transmitted into the network. These transmitted packets are sent to the echo/absorb board where they are absorbed.

For the measurement of user traffic statistics, user connections to PACNET are monitored. A packet switch traffic measuring instrument (TMI) was developed for this purpose. It consists of a custom board for an IBM personal computer (PC) (McCulloch 1986). This board was developed because no commercially available board has the required bandwidth. The board consists of an 8088 microprocessor, an Intel 8274 multiprotocol controller, a DMA controller, a counter/timer and 64K of RAM. The board runs as a co-processor to the IBM PC. The TMI can monitor circuits with transmission speeds up to 48,000 bps. It records the following information onto a floppy disk for latter analysis: 1) the arrival and departure times of packets, 2) the length of each packet, 3) the header for each packet, 4) circuit status information.

8.3.2 Proposed Traffic Model

Early terminal/computer data traffic was modelled as purely enquiry/response traffic (Jackson and Stubbs 1969; Dudick *et al.* 1971). This is traffic where a user types a request to the computer (user input time) and after some interval (computer reaction time) the computer responds (computer output time). The user thinks about the response (think time) and then types another enquiry. A comprehensive study of this class of traffic was completed by Jackson and Stubbs (1969). Since then, the advent of type-ahead buffering (which allows a user to continue typing characters before the computer has responded to a previous set of characters), intelligent terminals (e.g. Videotex, electronic funds transfer at point of sell (EFT POS) terminals) and sophisticated application programs, make the underlying enquiry/response traffic assumption invalid. Type-ahead buffering complicates a traffic model since it is no longer possible (or appropriate) to identify the end of one enquiry/response and the beginning of the next. Intelligent terminals are those that interpret user characters and communicate with the computer only when required. The literature contains little on these traffic processes (Pawlita 1981), which are only loosely related to the user processes. The development of a general traffic model which is capable of characterising such traffic may be impossible. It may be more appropriate to develop specific models for each type of traffic process.

Figure 8.8 shows a short typical time sequence of the traffic on connections A and B. It is immediately apparent from figure 8.8 that the traffic is bursty (i.e. packets do not arrive completely at random but are clustered together into bursts). There appears to be little relationship between packet arrivals in one direction and arrivals in the other direction (i.e. it is not possible to identify think/computer reaction times; *cf.* Jackson and Stubbs 1969), apart from the fact that arrivals in both directions are clustered together.

The traffic model proposed in this subsection is general enough to describe all four traffic generation processes measured. The traffic generation process for each logical channel can be broken into bursts, each burst containing a set of n packets. The set of n packets is considered to form a message. The definition of the random variables used to describe the packet generation process are:

- P_i = interarrival time between data packets on the same logical channel within a burst.

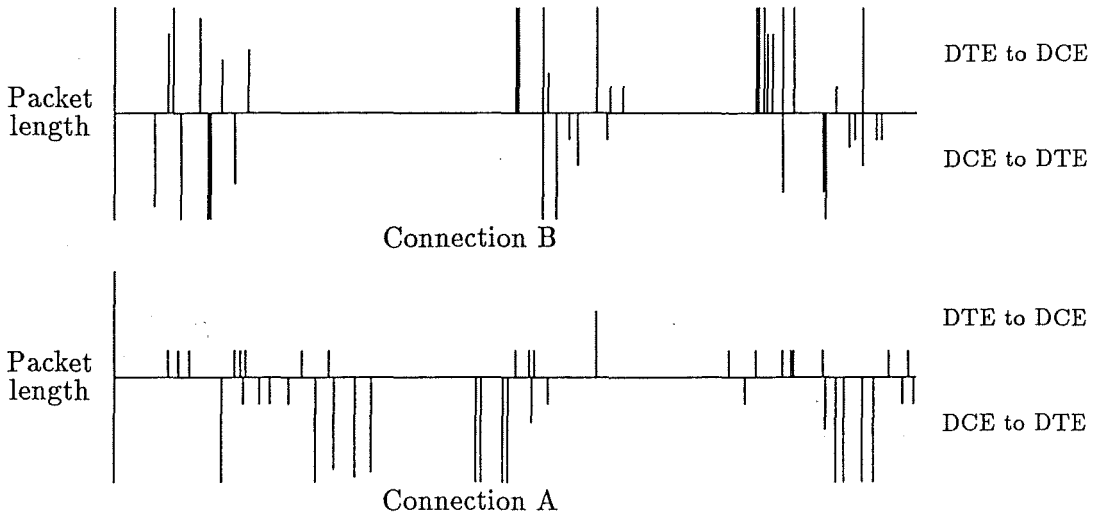


Figure 8.8: A short typical time sequence of the traffic on connections A and B. The central horizontal line represents the time axis, and the positions of vertical lines represent the times of arrival of data packets. The length of each vertical line is proportional to the length of the corresponding packet, and lines above the time axis represent packet flow from the DTE (user) to the DCE (network), and vice versa.

- W_i = time during P_i when the sender could not transmit a packet because of flow control.
- I_i = time during P_i when a data packet on a different logical channel is being transmitted and not overlapping W_i.
- M_j = interarrival time between message on the same logical channel.
- N_j = number of packets in the j^{th} message.

These definitions are explained pictorially in figure 8.9.

A packet belongs to a burst provided its interarrival time is less than the maximum allowed interpacket gap (MAIG) (Jain and Pouther 1986). The MAIG is obtained from the breakpoint (i.e. the intersection of two straight lines fitted to a logarithmic plot) constructed by plotting the logarithmic scale of relative frequency of packet interarrival time against packet interarrival time, for interarrival times up to approximately five seconds. The reasoning behind this is that a batch Poisson arrival process is assumed, and the increased arrival rate for short interarrival times is due to arrivals within bursts (Jain and Pouther 1986).

8.3.3 Traffic Measurements

Information about the packets that flowed on each of the four user connections were recorded for one week (five working days), and the recording process was conducted over four consecutive weeks (from Monday 2nd March 1987 until Friday 27th March 1987). A number of measurements concerning the packet flow for each user was

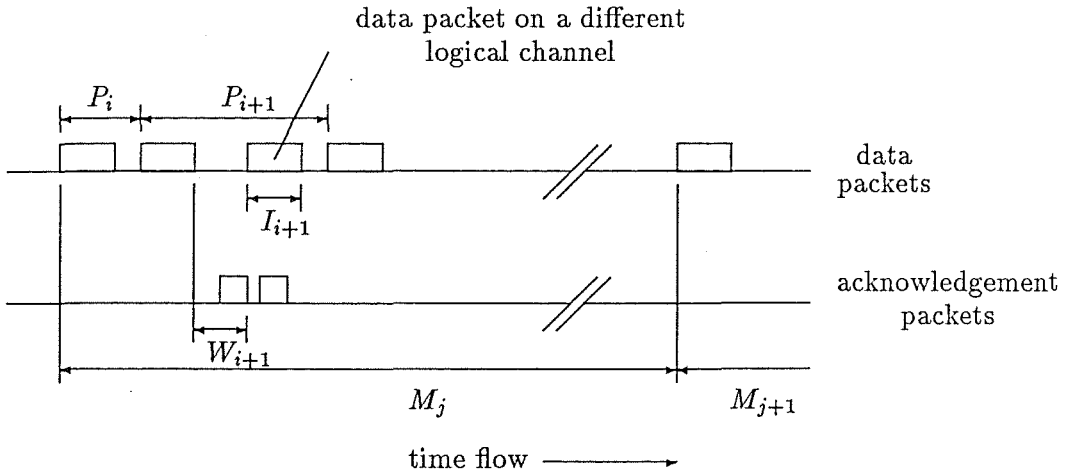


Figure 8.9: Pictorial explanation of the random variables defined in the traffic model introduced in §8.3.2.

extracted from the recorded information. This subsection presents the result of each measurement on a histogram. Each histogram is formed by averaging the result of each measurement over the five recording days. Each users circuit is composed of two channels (i.e. packet flow in the DTE to DCE direction and vice versa), therefore there are eight sets of results for each measurement: one for each user, and one for each direction of transmission (i.e. each channel). The results of the same measurement but for a different user or channel are presented on a different histogram. Histograms for the same measurement are grouped together into the same figure. Each figure, except one, consists of eight histograms and is split into two groups of four. The top and bottom group of four histograms in each figure present the results for traffic flow in the DCE to DTE, and DTE to DCE directions respectively. The results for connections A,B,C and D are presented in the histogram at the top left hand corner, top right hand corner, bottom left hand corner and bottom right hand corner of each group of four histograms respectively.

Figure 8.10 shows histograms obtained by plotting the logarithm of the relative frequency of occurrence of packet interarrival time against packet interarrival time, for interarrival time up to 5 seconds. In each case the histogram can be characterised by two straight lines. The intersection point of these straight lines establishes the position of the breakpoint (refer to §8.3.2). The packet interarrival time at which the breakpoint occurs specifies the maximum allowed interpacket gap (MAIG). For packet flow from the DCE to the DTE, the MAIG for each user is about 2 seconds. For packet flow in the other direction (i.e. from the DTE to the DCE) the MAIG for each user is about 1 second.

Figure 8.11 shows histograms obtained by plotting the logarithm of the relative frequency of occurrence of message interarrival time against message interarrival time, for interarrival time up to 60 seconds. In each case a straight line can be reasonably accurately fitted to the histograms, suggesting that message interarrival time are exponentially distributed.

Figure 8.12 shows histograms obtained by plotting the frequency of occurrence of packet interarrival times within packet bursts against packet interarrival times

within bursts. These interarrival times are clearly not exponentially distributed. Two distinct peaks occur for packet flow from the DCE to the DTE. This does not occur for packet flow in the other direction. Packet interdeparture times from PACNET seem to fall into two time intervals (i.e. quantised into two intervals). This quantised effect seems to be related to the time PACNET requires to service a packet, which is approximately constant for each packet. If there is a queue of packets waiting for service within PACNET then the stream of packets departing PACNET must be separated in time by the time required to service each packet (i.e. interdeparture times are approximately equal to the time required to service a packet). If a source user sends a burst of packets to a destination user, these packets queue for service within PACNET. If the interdeparture times of packets to the destination are measured, a peak is expected to occur at times (and multiples thereof) equal to the time it takes PACNET to service a packet. The first peak in figure 8.12 (i.e. the peak that occurs for the shorter interdeparture time) occurs at approximately 200ms, and is of similar duration to the time PACNET requires to service a packet. A possible explanation for the second peak (the peak at twice the interdeparture time of the first) is that PACNET services a packet for a different user before serving the next packet for the destination user being measured. More study is required into this phenomenon before a more definite conclusion can be reached. However, it does show that packets departing PACNET are correlated.

Figure 8.13 displays the autocorrelation $R(n)$ of packet and message interarrival times. The left and right hand column of graphs give the autocorrelation of packet interarrivals from the DCE to DTE and DTE to DCE respectively. Plotted on each graph are four curves, with each corresponding to one of the four circuits. The top pair of graphs show the autocorrelation of packet interarrival times for interarrivals up to 60 seconds. The centre pair of graphs gives the correlation between packet interarrival times within a packet burst. The bottom pair of graphs gives the autocorrelation of message interarrival times. The bottom pair indicate minimal correlation of message interarrivals. This together with the approximate straight lines of the curves shown in figure 8.11 strongly suggest message interarrivals can be characterised by a Poisson process (Jain and Pouther 1986). The centre pair of graphs indicate correlation between packets in a burst as was speculated in the previous paragraph.

Figure 8.14 shows histograms obtained by plotting the frequency of occurrence of packet length within packet bursts against packet length, where the last packet in a burst is excluded. The maximum length packet that PACNET accepts is 135 bytes. This is made up of seven bytes of network required information (which is contained in the header) and 128 bytes of user data. The histograms in figure 8.14 have peaks at packet lengths of 64 and 128 bytes. Telecom charges users according to packet length. There are two packet length charges. One charge for packets containing 64 bytes of user data or less, and twice this charge for packet lengths containing more than 64 bytes of user data. Thus, the most cost efficient method for transmitting packets is to fill packets with either 64 or 128 bytes. Figure 8.14 confirms that users do this.

Figure 8.15 shows the histograms obtained by plotting the frequency of occurrence of the length of the last packet within packet bursts against packet length. Figure 8.15 shows that the last packet in a burst contains packets of all lengths. This is what is to be expected if a packet burst is considered to be a complete mes-

sage, where the user splits a message into either full or half length packets, with the remaining piece of the message being sent in the last packet of a burst.

Figure 8.16 shows the histograms obtained by plotting the frequency of occurrence of the number of packets which occur within packet bursts against number of packets. The histogram is different for each connection and direction of packet flow. For circuit A there is an order of magnitude more packets sent from the host computer than is sent to the host computer. For connections B,C,D more than twice as many packets are sent from the host computer than are sent to the host computer. Consequently, the number of packets within a burst sent to the host computer is significantly smaller than that sent from the host computer. No particular probability distribution seems appropriate to characterise the number of packets within a burst. One histogram has at least two peaks, while the other histograms have a single peak and appear to be geometrically distributed.

The amount of time each DCE is unable to send packets because of network blocking is quite small when compared to packet interarrival times. When the network blocking time W_i is subtracted from the packet interarrival times, the histograms exhibit little in the way of differences. This is also true when the interference traffic time I_i is subtracted from the packet interarrival times.

These preliminary traffic results indicate that a batch Poisson traffic arrival process seems an appropriate model for at least the four connections examined (*cf.* Jain and Pouther 1986). Most analytic packet switching models assume a Poisson packet arrival process. However, it is well known that a Poisson approximation to a batch Poisson process seriously underestimates the required buffer and link resources required in a packet switching network (Chu 1970).

Analytic models are at present woefully inadequate for predicting the performance of real-world communication networks. Computer simulations will remain a cornerstone for analysing communication networks at least into the immediate future. Comprehensive simulation studies investigating how user traffic statistics affect the amount of resources required in a network for a given GOS is lacking. Such a knowledge together with accurate user traffic statistics (particularly when it is considered that many new services with unknown traffic statistics are being connected to packet switching networks) are necessary to manage a network well. The development of a suitable traffic measuring equipment is required if accurate traffic measurements are to be made. The traffic measuring instrument (TMI) described in §8.3.1 provides the basic facilities needed for traffic measurement, and gives a useful basis for the development of more elaborate instruments incorporating graphical display of real time network performance and traffic statistics.

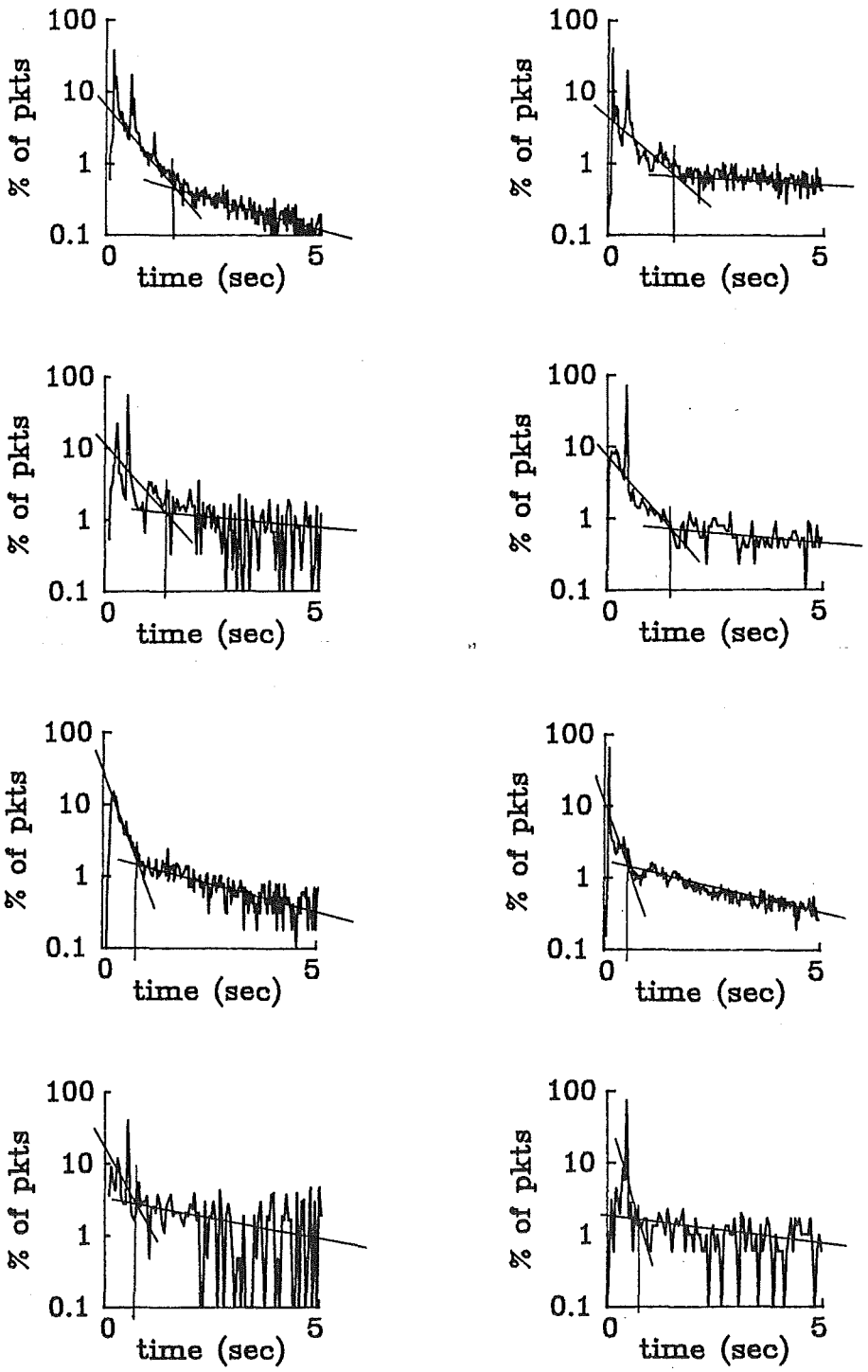


Figure 8.10: Histograms obtained by plotting the logarithm of the relative frequency of occurrence of packet interarrival time against packet interarrival time, for interarrival times up to 5 seconds. Top group of four histograms present the results for the DCE to DTE direction. Bottom group of four histograms present the results for the DTE to DCE direction (refer to first paragraph in §8.3.3).

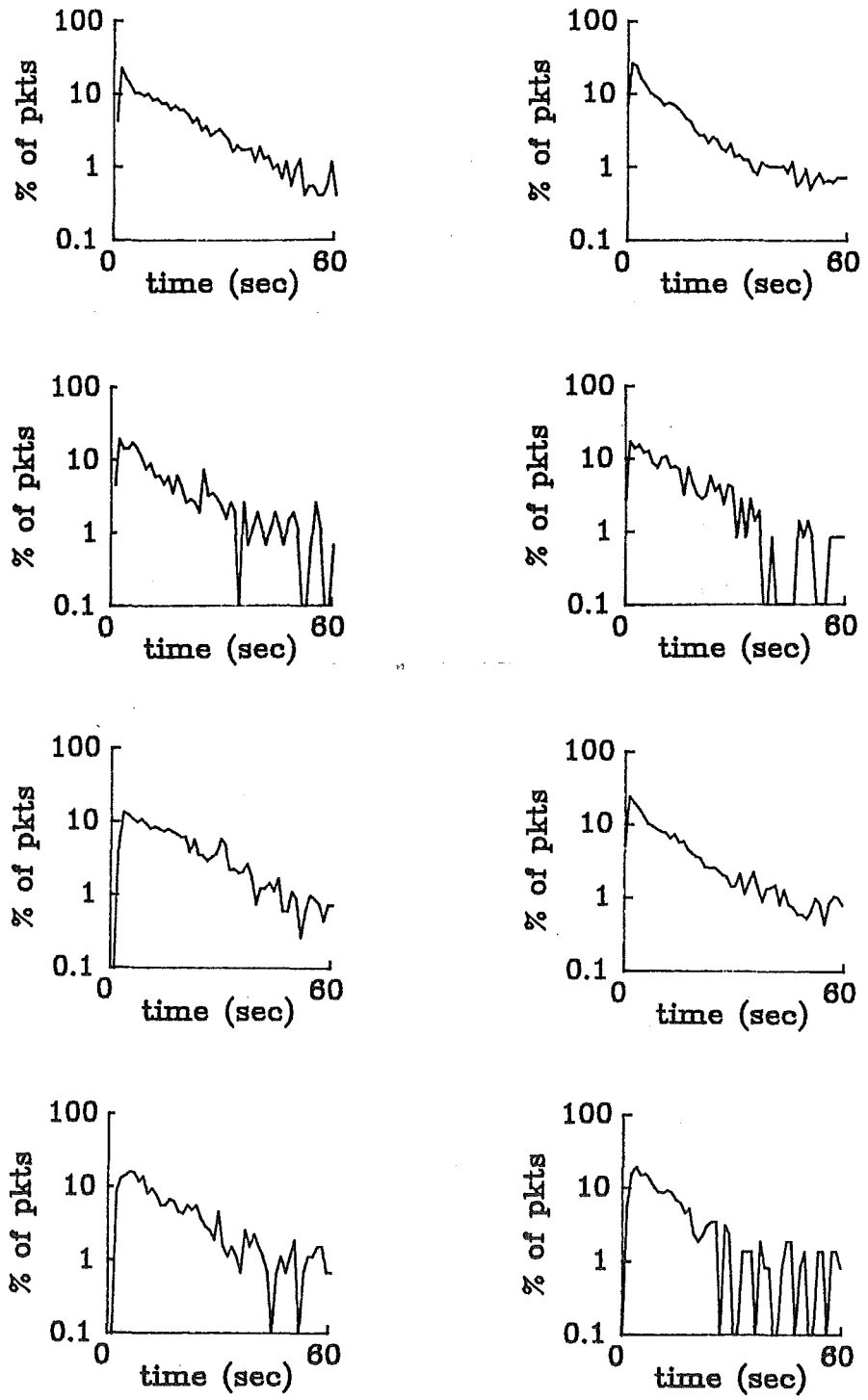


Figure 8.11: Histograms obtained by plotting the logarithm of the relative frequency of occurrence of message interarrival time against message interarrival time, for interarrival times up to 60 seconds. Top group of four histograms present the results for the DCE to DTE direction. Bottom group of four histograms present the results for the DTE to DCE direction (refer to first paragraph in §8.3.3).

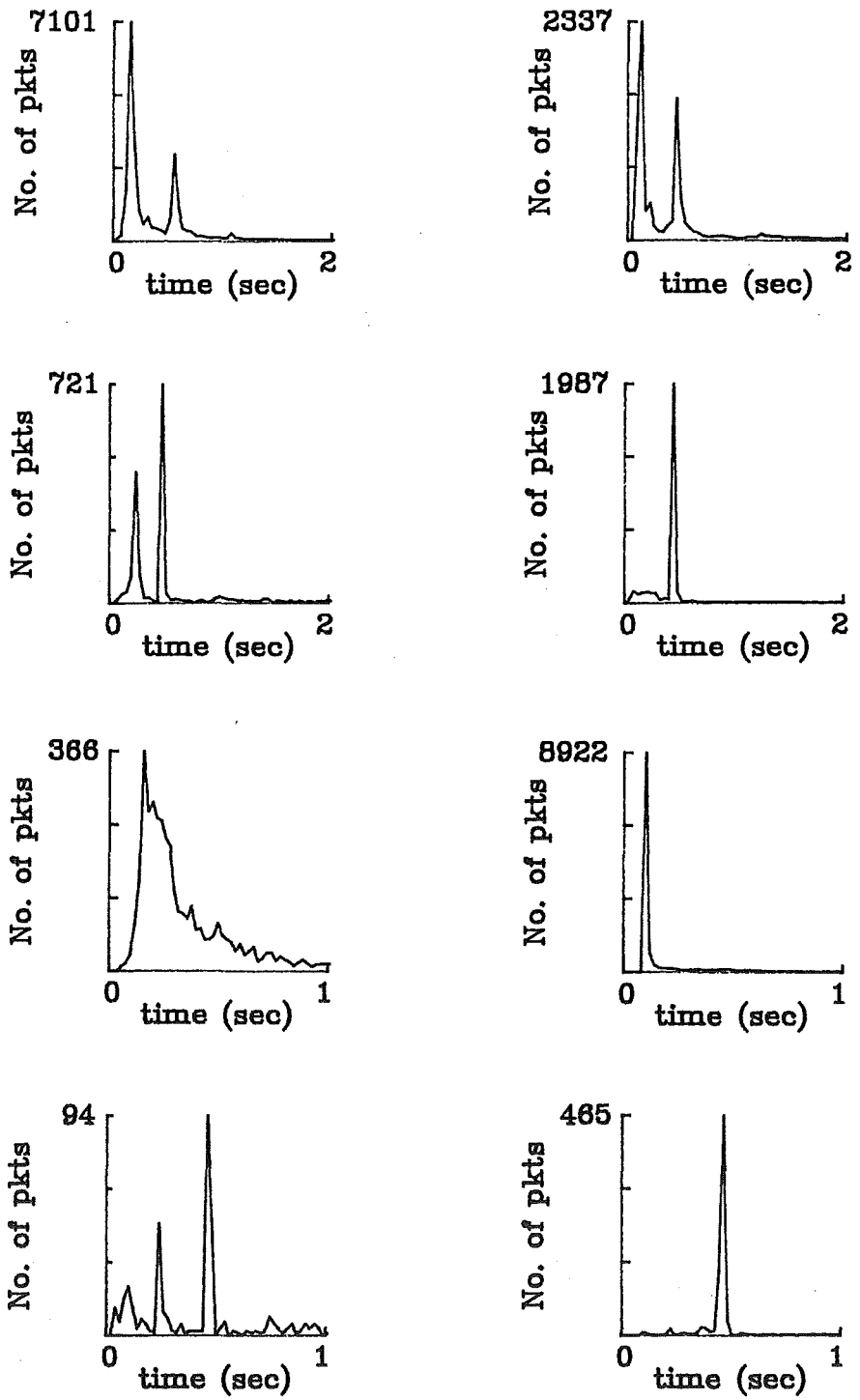


Figure 8.12: Histograms obtained by plotting the frequency of occurrence of packet interarrival time within packet bursts against packet interarrival time within bursts. Top group of four histograms present the results for the DCE to DTE direction. Bottom group of four histograms present the results for the DTE to DCE direction (refer to first paragraph in §8.3.3).

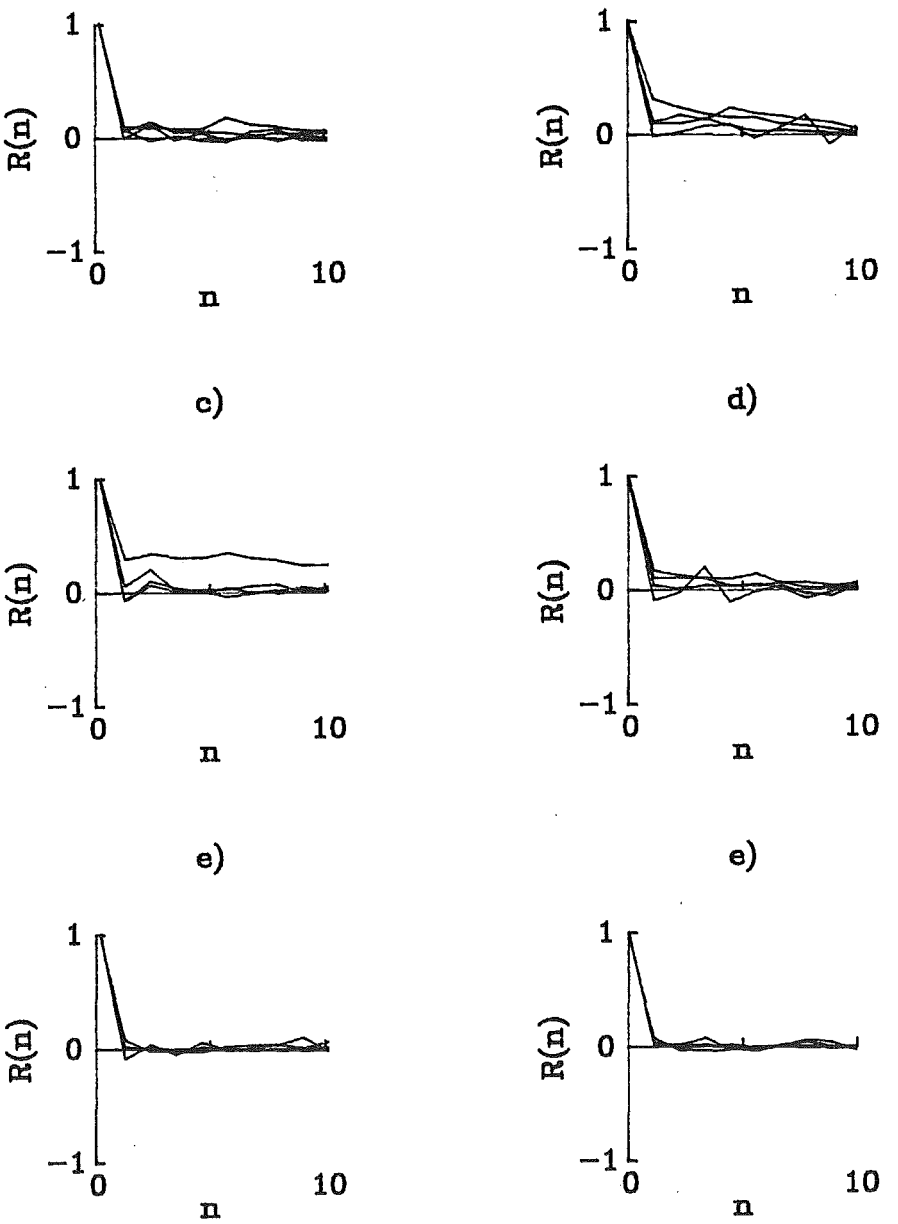


Figure 8.13: Graphs showing the autocorrelation $R(n)$ between packet and message interarrival times. The left and right hand column of graphs gives the autocorrelation between packet interarrival times for the DCE to DTE and DTE to DCE directions respectively. Plotted on each graph is four curves, each curve corresponds to one of the four circuits monitored.

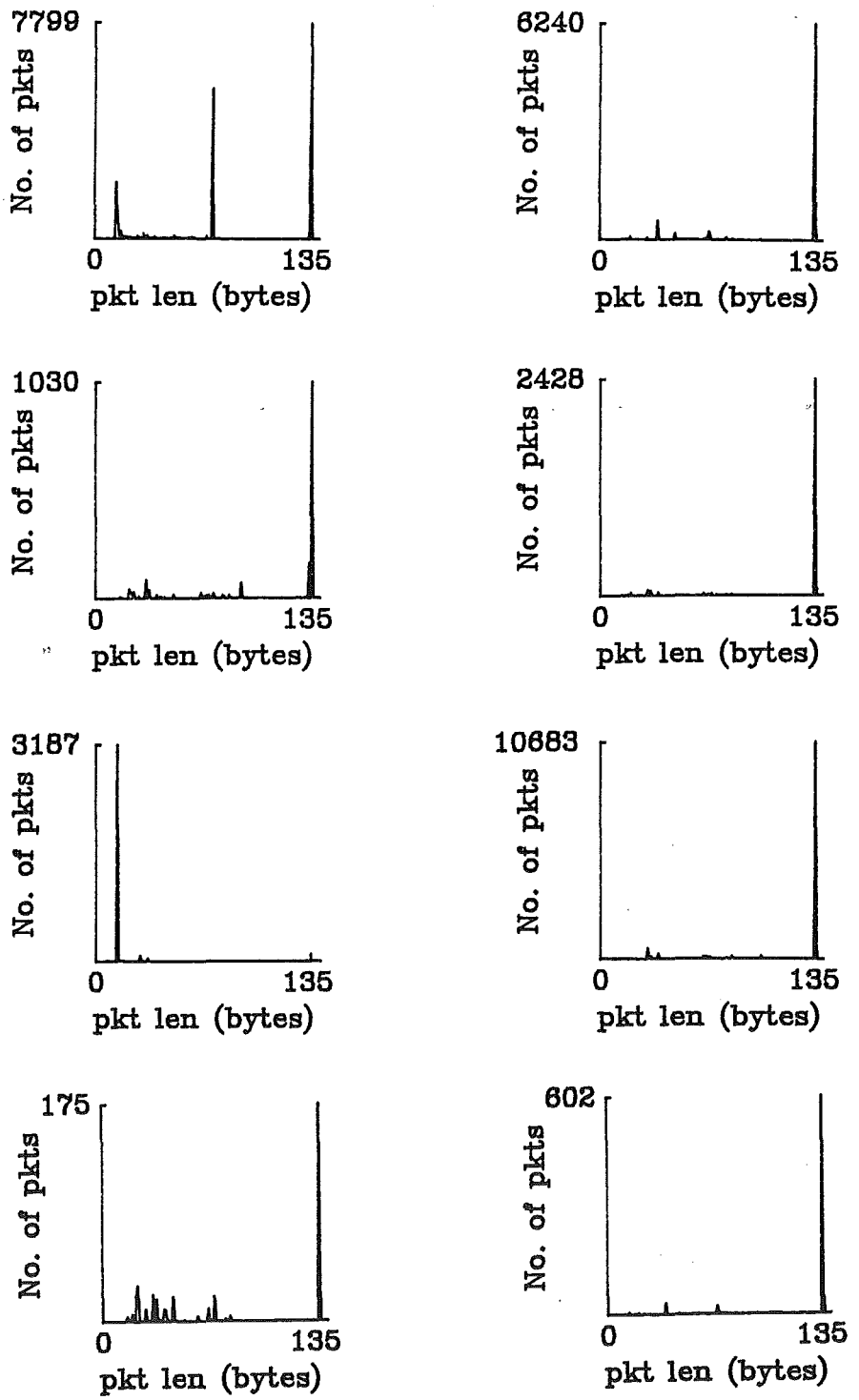


Figure 8.14: Histograms obtained by plotting the frequency of occurrence of packet length within packet bursts against packet length, with the last packet in the burst excluded. Top group of four histograms present the results for the DCE to DTE direction. Bottom group of four histograms present the results for the DTE to DCE direction (refer to first paragraph in §8.3.3).

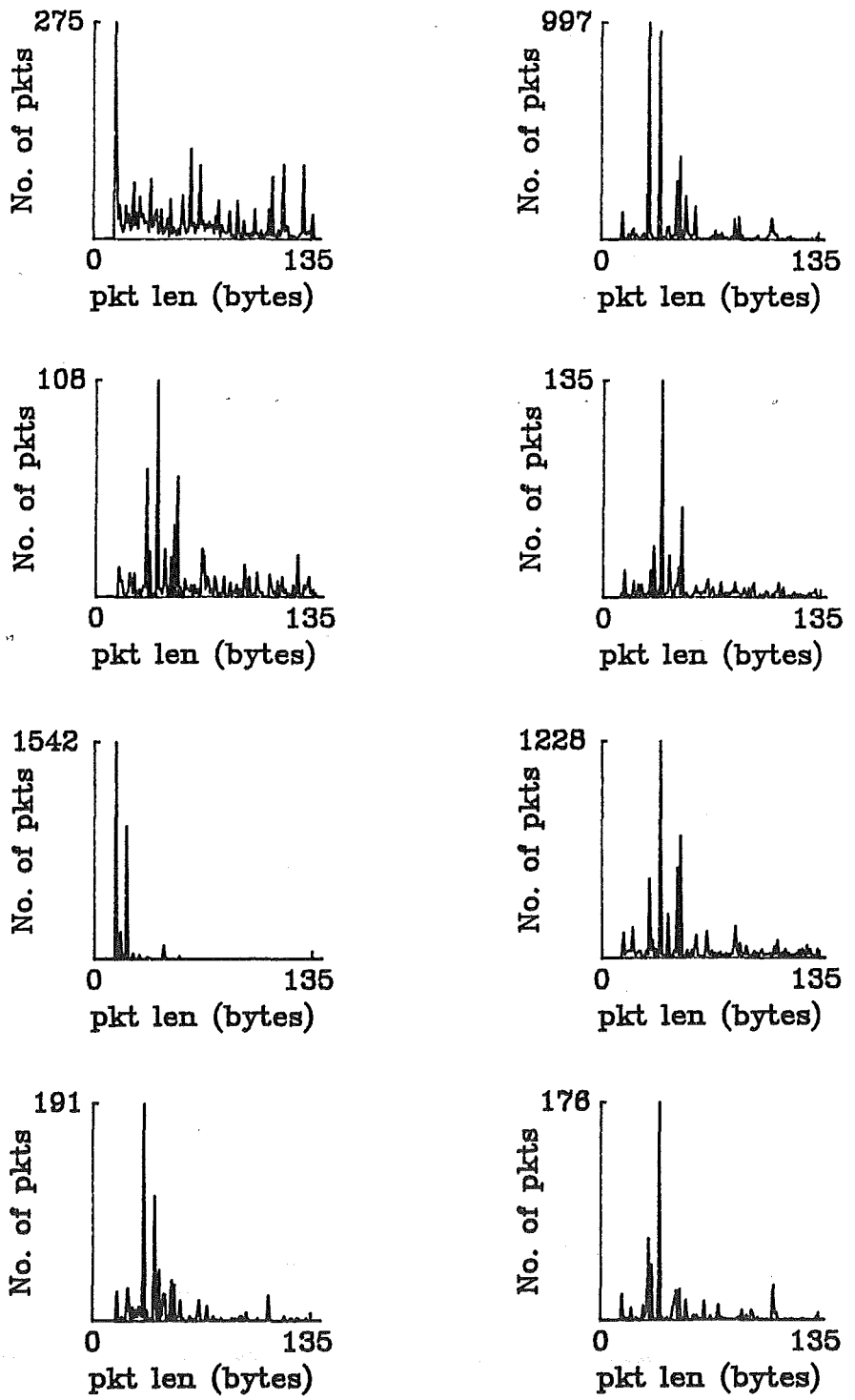


Figure 8.15: Histograms obtained by plotting the frequency of occurrence of the length of the last packet within packet bursts against packet length. Top group of four histograms present the results for the DCE to DTE direction. Bottom group of four histograms present the results for the DTE to DCE direction (refer to first paragraph in §8.3.3).

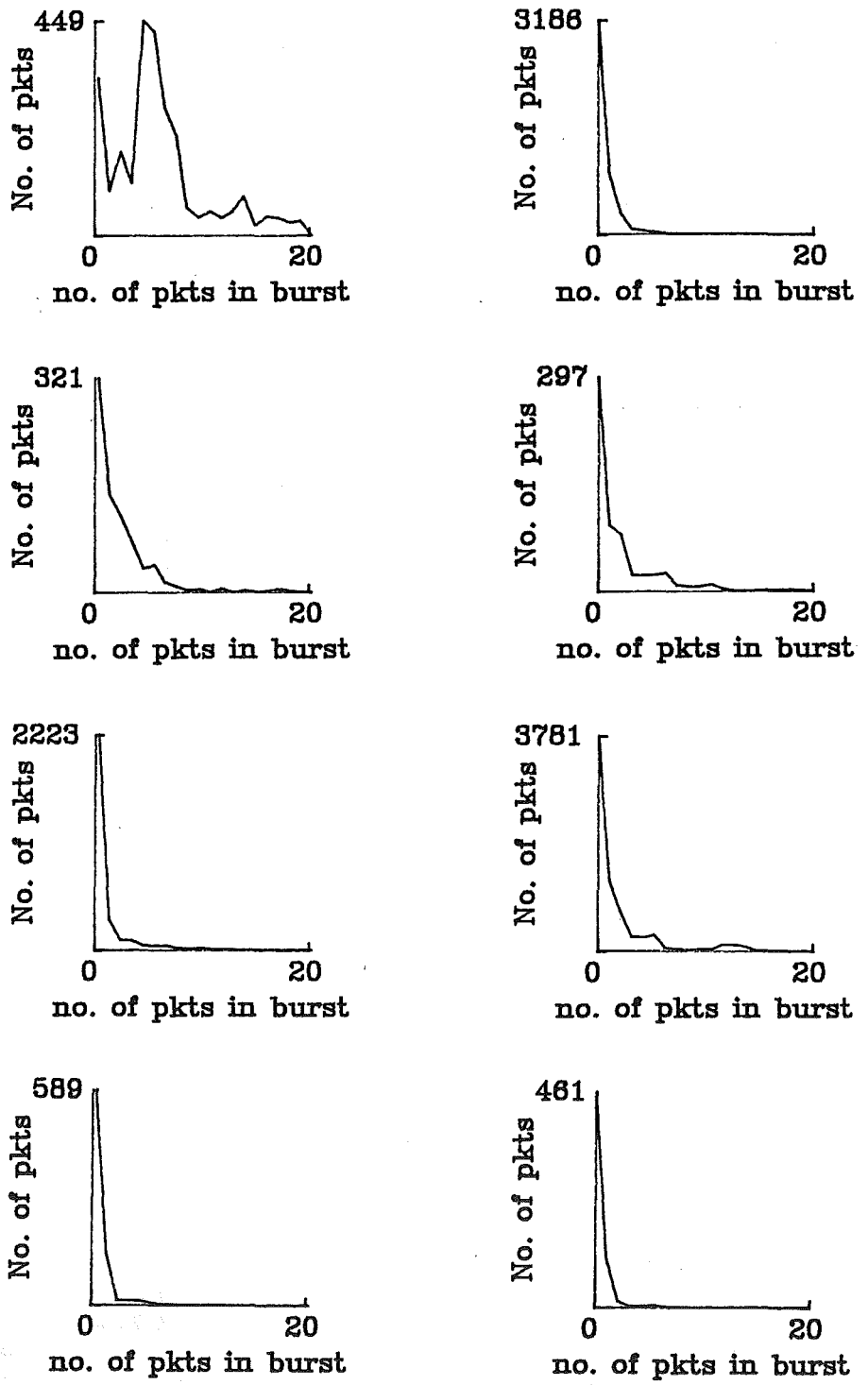


Figure 8.16: Histograms obtained by plotting the frequency of occurrence of the number of packets which occur within packet bursts against number of packets. Top group of four histograms present the results for the DCE to DTE direction. Bottom group of four histograms present the results for the DTE to DCE direction (refer to first paragraph in §8.3.3).

Chapter 9

Conclusions and Suggestions for Further Work

The theme of this thesis has been the investigation of the potential applications, consequences and prevention of deterministic chaos in technologically important areas. Four such areas have been investigated; electrical noise, electrical signal sources, data encryption and packet switching.

Chapter 1 reviews deterministic chaos in both dissipative and conservative systems. It has three goals: 1) to give an intuitive feel for how complicated behaviour can arise in deterministic systems, 2) to introduce many of the terms used in the specialist literature, 3) to provide the theoretical basis for the later chapters. Many of the general mathematical terms and concepts required to provide the proper setting for explaining deterministic chaos are introduced in chapter 2. Chapter 3 discusses some of the philosophical and practical consequences that stem from deterministic chaos. Chapter 4 is devoted to the historical development that has led to the present renewal of interest in dynamical systems.

This chapter summarises the important aspects and draws conclusions from the results reported in chapters 5,6,7 and 8. It also includes suggestions for further research. The conclusions and suggestions for each chapter are collected into separate sections.

9.1 Generating Deterministic Noise

Chapter 5 introduces a computational framework (readily implementable in hardware, because of the simplicity of its components (Smith 1987)) seemingly capable of generating sequences of numbers exhibiting arbitrary probability density functions and power spectra. The integral equation (5.17), developed in §5.2, permits a wide class of pdfs to be predicted or synthesised to a useful level of accuracy. The practical value of the theoretical/computational approach introduced in §5.3 and §5.4 would be significantly enhanced if complementary methods for predicting/synthesising power spectra could be devised. This might permit a simple and versatile coloured noise generator (Bates and Murch 1987) to be implemented with some ease. However, such a noise generator could, as things presently stand, be implemented readily enough with the aid of a lookup table derived from extensive numerical investigation of (5.22)

in §5.3.

The most immediately striking aspect of the concept of a hierarchy of recursive loops is that it constitutes such an elementary means of generating a noisy process having an excess low frequency character. It might be conjectured that the demonstrated (in §5.3) effect of the nonlinearity on the power spectrum of a variable-gain sequence constitutes a significant addition to the literature on 1/f noise, because its very simplicity may suggest a hitherto overlooked physical explanation for the wide occurrence of such noise. The connection of recursive loops into hierarchies constitutes but one connection structure. Other structures and/or recursive loops might prove more versatile for generating sequences with arbitrary pdfs and power spectra (PS). To permit the formulation of connection structures and recursive loops which are optimally suited to generating arbitrary pdfs and PS a more complete mathematical framework is needed. This framework might take the form of an iterative optimisation procedure. Recursive loop nonlinearities and connection structure weightings (and/or other parameters characterising the connection structure) are adjusted such that the pdf and PS (or parameters characterising these) of the sequences generated by the structure are optimised according to some criterion (e.g. to enable the pdf to be altered without affecting the PS and vice versa). Such a framework requires the development of an algorithm characterising the way recursive loop nonlinearities and connection structure weightings should be altered at each iteration.

The final paragraphs of §5.4 may have some biological implications. It is interesting to note that chaotic behaviour may be an inherent and necessary element in the healthy formation and functioning of living organisms (Sporns *et al.* 1987; Goldberger and Rigney 1988). For example, Goldberger (1988) argues that healthy human hearts exhibit chaotic dynamics, while the dynamics leading to sudden cardiac arrest are periodic and not chaotic. This suggests that chaos might be a crucial universal characteristic of biological systems, thereby providing yet another fruitful area of collaboration between systems engineers and life scientists. Many biological rhythms can be upset by alterations to physiological parameters. Such alterations may be able to be modelled by changes to gain settings and/or nonlinearities of recursive loops arranged into hierarchies. It is possible that insights may thereby be gained into the functioning of biological systems. This approach might be appropriate for modelling the global characteristics of biological processes which are so intricate that it is impracticable to model them more fully.

The final section of chapter 5 invokes Lyapunov exponents and information theory to shed light on variable-gain sequences. Figure 5.13 shows that mutual information (i.e. the degree to which the future can be predicted) increases with increasing noise amplitude (i.e. for $\alpha \rightarrow 0$), and in spite of the increasing Lyapunov exponent (both conditions, increasing noise amplitude and increasing Lyapunov exponents, generally reduce the predictability in deterministic systems). It has been conjectured that nonuniformity is a necessary condition for the occurrence of noise-induced predictability (*cf.* Herzel and Pompe 1987). The results presented in §5.5 provide further evidence in support of this conjecture. The significance of these results is that, for some chaotic dynamical systems, increasing the amplitude of additive noise actually increases the predictability of the system behaviour (i.e. reduces future uncertainty). There are, however, many questions which remain open, such as: does noise-induced predictability occur in the majority of systems? To what extent does noise-induced predictability increase predictability? How should the noise be introduced? Although

noise is already sometimes introduced into control systems (to reduce stiction and other undesirable aspects of mechanical activators), additive noise has hitherto not been considered as a means of improving the predictability of system behaviour. Noise-induced predictability may be of technological importance since it might allow systems and processes, which are at present unpredictable (e.g. system reliability, weather), to be made (more) predictable and useful.

9.2 Sinusoidal Oscillator Phase Noise

Increasing demand on the frequency spectrum required for communications (Rutman 1978; Robins 1984) and on the accuracy of time and frequency standards (Jespersen *et al.* 1972; Rutman 1978) has continually reduced the permissible level of the output noise from sinusoidal oscillators (i.e. oscillators with improved short and long term stability). In practice, spurious phase noise is of more concern than spurious amplitude noise because it is usually of a much higher level (Robins 1984, page 47). Chapter 6 assesses to what extent the phase noise in the output of a high quality sinusoidal oscillator can be attributable to deterministic chaos.

The noise performance of a typical real-world sinusoidal oscillator seems to be somewhat worse than established approaches to noise analysis of theoretical models of such oscillators would suggest (Rutman 1978) (Robins 1984, page 63). Two effects which tend to be neglected in such analysis are the signal (and signal-dependent) delay around the oscillator loop, and signal-dependent parameter variations. In practice, transit times of charge carriers across active components constitute most of the delay around an oscillator loop. Furthermore, stray capacitances within amplifiers are the parameters whose values are usually most dependent on the oscillator signal level. These effects introduce erratic intermodulation of the oscillator signal with itself. This causes noise which is already present and which is spectrally distributed around the carrier frequency (e.g. 1/f noise modulated onto the carrier, thermal noise), to become distributed over a wider spectral range (i.e. reduces the spectral purity of the oscillator signal). Also, intermodulation of the oscillator signal with itself causes the level of the harmonics of the carrier frequency to increase.

Two nonlinear gain oscillators, whose amplifier characteristics correspond, respectively, to a soft limiter and a tunnel diode are examined in §6.4.1 and §6.4.2. Since both amplifier characteristics are odd symmetric, the nonlinearity introduces odd harmonics (mostly third harmonic) into the oscillator signal. This is illustrated in figure 6.8(c) and figure 6.10(c) where the carrier to noise and carrier to third harmonic ratio tend to become comparable as the gain constant g is increased, while the carrier to second harmonic ratio remains very large. For the tunnel diode characteristic, the oscillator signal is chaotic for gains greater than 4.0. The carrier to noise ratio (plotted in figure 6.8(c)) rapidly degrades for gains above 4.0, while the carrier to third harmonic and carrier to second harmonic ratios are not significantly affected.

The results presented in figure 6.8(d)-(f) and figure 6.10(d)-(f) show that increasing the level of signal-dependent variation in capacitance (i.e. increasing δC) degrades the carrier to second harmonic ratio significantly, but leaves the carrier to third harmonic unaffected. The results presented in figure 6.10(g)-(h) and figure 6.8(h)-(i) show that increasing the level of signal-dependent variation in delay (i.e. increasing

δt_d) degrades the carrier to second harmonic ratio at an even faster rate (than when increasing the level of signal-dependent variation in capacitance), while the carrier to third harmonic ratio actually decreases with increasing δt_d. Even quite small values of δt_d or δC can degrade the spectral purity by significant amounts. For low values of the gain constant g the effect of δt_d and δC on the spectral purity is less significant than for larger values of g.

Signal-dependent variations in circuit capacitance δC and delay δt_d cause the level of the second harmonic of the carrier to increase. This occurs because such signal-dependent variations do not distort the positive and negative cycles of the oscillator waveform identically (i.e. the distortion does not cause the waveform to possess odd symmetry). Such distortion introduces even (mostly second) harmonics of the carrier into the oscillator waveform. Signal-dependent variations in circuit capacitance and delay have less effect on the level of the third harmonic. Such variations increase the intermodulation of the oscillator waveform (i.e. they alter the shape of the oscillator waveform by introducing harmonics of the carrier) but does not introduce noise with a continuous frequency spectrum. However, if the parameters characterising an oscillator are set such that its operation is close to being chaotic, signal-dependent delay and capacitance may cause the oscillator signal to alternately start and stop being chaotic, resulting in apparently random behaviour. While such a condition can tip the oscillator into and out of chaos, it is not the fundamental cause of it. This implies that signal-dependent variations in circuit parameters operating in conjunction with other mechanisms within the oscillator loop (which on their own may not induce chaotic behaviour) can together induce chaotic behaviour (i.e. generate apparent noise with a continuous frequency spectrum).

The nonlinear gain oscillator whose amplifier is a tunnel diode generates an oscillator signal which is chaotic when a fixed arbitrary length signal delay is introduced into the oscillator loop and the gain constant g is sufficiently large (note that Kitano *et al.* (1983) have examined a similar but different oscillatory system). Although introducing signal-dependent delay and capacitance widens the spectral distribution of the chaotic oscillator signal, the widening appears to be quite limited. Computational experience indicates that the spectral distribution about the carrier widens by less than 10% for a δt_d or δC of 0.1 when $g = 5$. Quantifying how the spectral distribution of a chaotic oscillator signal (and also of additive natural noise when it is introduced into an oscillatory loop) depends on the level of signal-dependent delay and capacitance represents a useful area for future study.

Conditions sufficient for a nonlinear gain oscillator to exhibit deterministic chaos are, first, the existence of a signal delay (it is not necessary for the delay to be signal-dependent) around the oscillator loop and, second, certain types of amplifier nonlinearity. It is assumed by many researchers (*cf.* Beurle 1956; Hafner 1966; Robins 1984) that the amplifier nonlinearity is a smooth monotonically increasing function (e.g. soft limiter). However, in reality this assumption may be false. In fact, amplifier nonlinearity is more likely to be bumpy (if for no other reason than that almost all one-dimensional functions form recursive loops that can be chaotic; refer to §1.1.6). A bumpy characteristic is modelled in §6.4.3 by adding small sinusoidal bumps to a soft limiter characteristic. As is shown in the right hand graph in figure 6.12 the size of the bumps determines the amplitude of the resulting chaos. It seems inescapable, therefore, that if can amplifier characteristic exhibits bumps then any oscillator in which the amplifier is incorporated must exhibit deterministic chaos. Mechanisms

that might cause such a bumpyness in the case of the field effect transistor (FET) are postulated in §6.4.3. Note that parts of the bumpy nonlinearity need to possess negative curvature if chaos is to exist. However, the existence of such bumps in practice has yet to be confirmed. The literature reveals nothing about the existence of such bumps. This may indicate that bumps, if they exist at all, are so small that they are difficult (or nearly impossible) to experimentally detect and/or the existence of such bumps have been hitherto considered unimportant for inclusion in oscillator models. The existence of bumps is of considerable potential practical significance since they may ultimately determine the lower limit to oscillator noise. The development of (or extension of existing) theoretical models incorporating such bumps, and undertaking experiments capable of detecting them in semiconductor devices, represent useful areas for future research.

A linear gain controlled oscillator is examined in §6.5. The level of the second and third harmonics of the carrier are significantly lower than for the nonlinear gain oscillators. Since the linear gain controlled oscillator incorporates automatic gain control (agc), if the loss around the loop should change due to variations in circuit capacitance or signal delay, the agc can counter this by adjusting the amplifier gain. The results presented in figure 6.13(b) and figure 6.13(c) show that increasing δC or δt_d degrades the carrier to second harmonic ratio significantly, but leaves the third harmonic largely unaffected. Although the spectral purity does not degrade to the same extent with increasing δt_d or δC as for the nonlinear gain oscillators, even quite small variations in t_d or C can degrade the spectral purity significantly. For both figure 6.10(b) and figure 6.10(c) the level of the third harmonic is 100dB down on the carrier, which is considerably less than the levels for the nonlinear gain oscillators. The level of the third harmonic is lower because the amplifier, being linear, introduces little waveform distortion. However, any signal-dependent variation in circuit capacitance or signal delay raises the level of the second harmonic significantly. Overall, the spectral purity of the linear gain controlled oscillator is considerably better than that of the nonlinear gain oscillators. Quantifying how the spectral distribution of the oscillator signal depends on the level of signal-dependent delay and capacitance, when additive natural noise is introduced into the oscillator loop, is a useful area for future work. For instance, it would be interesting to see if the spectral purity of the linear gain controlled oscillator is significantly better than that of the nonlinear gain oscillator when additive natural noise is introduced into the oscillatory loop.

The first report of complicated dynamical behaviour from a circuit first suggested by Chua (which now bears his name) was by Matsumoto (1984). Chua's circuit (also known as Chua's oscillator) has generated much interest for the following reasons:

- It is the simplest autonomous electrical circuit which can become chaotic.
- It is the only physical system that has been shown to be chaotic through, not only computer simulation (Matsumoto 1984), but also laboratory experiments (Zhong and Ayrom 1985) and mathematical analysis (Chua *et al.* 1986).
- It exhibits an immense variety of dynamical phenomena, including many typical 'bifurcations' and 'scenarios preceding chaos' (Matsumoto *et al.* 1986b).

The robustness (of the deterministic chaos generated by Chua's circuit) to circuit component values and to signal-dependent variations in component values is assessed

in §6.6. Matsumoto *et al.* (1985) report that the chaotic behaviour persists for the range of circuit parameters listed in (6.44). However, Matsumoto *et al.* (1985) do not report on how sensitive the chaotic behaviour is to the parameters characterising the nonlinearity. §6.6 demonstrates that the strange attractor of Chua's oscillator persists for the range of nonlinearity parameter values listed in (6.46).

§6.6 demonstrates that the chaotic behaviour generated by Chua's circuit is quite sensitive to:

- The introduction of a signal delay into the nonlinearity.
- The smoothing out of the corners of the 5-segment piecewise-linear resistor by the (smooth) fitting of cubics.
- The slope of the middle segment of the 5-segment piecewise-linear resistor.

A delay of greater than 0.01 (i.e. greater than 0.3% of the fundamental period of the oscillator signal) reduces the behaviour of Chua's circuit to a stationary point (i.e. it inhibits the spontaneous appearance of chaotic noise). In §6.4.3 signal delay and particular nonlinearity shapes are shown to be sufficient conditions for the nonlinear oscillators studied in §6.4 to exhibit deterministic chaos. In the case of Chua's circuit it is the absence of signal delay which is necessary. This suggests that deterministic chaos is sensitive to the shape of nonlinearities and the presence or absence of signal delay (i.e. if chaos occurs in a system with signal delay then in general it is not chaotic when the signal delay is removed and vice versa).

Certain kinds of small perturbations to the 5-segment piecewise-linear resistor in Chua's circuit can inhibit chaotic behaviour. Therefore, the chaotic behaviour generated by Chua's circuit is not robust to all forms of small perturbations made to the nonlinear function. If this is true in general systems which exhibit chaotic behaviour (and in particular oscillators), then it may be possible by making relatively small system changes to inhibit the chaotic behaviour, when such behaviour is undesirable (as it is in high stability oscillators, for instance).

Understanding why Chua's circuit is sensitive to certain kinds of small circuit perturbations may provide further insight into chaotic dynamics in general. It is particularly surprising that the introduction of a short signal delay into the nonlinearity inhibits chaotic behaviour in Chua's circuit. Understanding why this occurs interests me personally and represents a useful area for further research. The sensitivity of the chaotic behaviour exhibited by Chua's circuit to certain small perturbations of the nonlinearity does not surprise me to the same extent. The extensive analysis presented by Chua *et al.* (1986) requires the nonlinearity to be piecewise-linear. Any smoothing out of the corners of the nonlinearity alters the fundamental character of Chua's circuit. Another useful area for further work would be to analyse Chua's circuit for a wide range of nonlinearities and attempt a comprehensive characterisation of the sensitivity of the occurrence of deterministic noise to various classes of nonlinearity.

9.3 Chaotic Data Encryption

Chapter 7 assesses to what extent the seemingly random numbers generated by chaotic dynamical systems are suitable for data encryption. Chaotic dynamical sys-

tems have properties which are highly desirable for data encryption. The specific use of chaotic dynamical systems for data encryption seems so far to have been overlooked. It is convenient to split data encryption using chaotic dynamical systems into two categories: isolated chaotic encryption (refer to §7.3) and influenced chaotic encryption (refer to §7.4). Isolated chaotic encryption is based around the one time tape (refer to §7.1), while influenced chaotic encryption is an apparently completely new method of encryption introduced in this thesis.

The most practical way to implement chaotic encryption is with the aid of a digital computer (e.g. microprocessor) and/or dedicated hardware. All digital computers represent real numbers to a finite precision, which has two consequences. First, only a finite number of real numbers can be represented. Second, there is a real number (called the machine epsilon) which is the smallest number that can be represented by the computer. If a process using infinite precision numbers is modelled on a computer, the resulting model unavoidably uses finite precision numbers. The longest sequence of numbers that can be generated without repetition is necessarily finite and is therefore not genuinely chaotic.

Iterating a chaotic map with finite precision numbers has two consequences. The first is that sequences repeat and the second is that there exist a finite number of different sequences which in general have different period lengths (some of the periods can be very short). If chaotic maps are to be useful for encryption, the following two conditions must be fulfilled:

- The precision of the numbers used to iterate a map must be sufficient to ensure that the periods of the sequences generated by the model of the chaotic process are long enough for the particular encryption application.
- Initial conditions that generate sequences with periods shorter than required for the particular encryption application are avoided.

Little appears to be known about exactly how computation using finite precision numbers effects the dynamics of a chaotic system (*cf.* McCauley 1988, Chapter 5; Lichtenberg and Lieberman 1983, page 276). If more were known about this, it might help one to decide which chaotic systems serve best for pseudo-random number generation. It is possible to check that the essential properties (i.e. pdf and power spectrum) of numerically generated solutions are valid. This can be achieved by comparing numerical solutions generated using increased numerical precision, and verifying that the solutions are converging to a particular solution (such checks are generally a part of any numerical study). For example, the power spectra shown in figure 7.4 are similar although the precision of the computations ranges from numbers having 7 decimal digits to 16 decimal digits. This is of practical importance, since it suggests that the large body of dynamical results obtained for infinite precision numbers also applies to dynamics generated by finite precision numbers.

As regards data encryption, the important aspects of the dynamics of a chaotic system when implemented on a computer are:

- The length of the period of the sequence (i.e. the number of times that the chaotic equations must be iterated before the sequence repeats).
- The amount of information the sequence contains (i.e. the entropy of the sequence).

Figure 7.3 shows that when the size of the quantisation interval q is made smaller than about 10^{-9}, the average length of the period p of the sequence increases with q approximately twice as fast as when q is greater than 10^{-9}. In any practical implementation, therefore, q should be less than 10^{-9}. A value for q of 10^{-16}, which can easily be obtained with existing hardware, gives (refer to §7.2 and figure 7.3) a value for p of roughly 10^{10}, which seems adequate for (at least) the great majority of encryption applications.

Isolated chaotic encryption utilises chaos to realise a pseudo-random number generator, which supplies pseudo-random numbers to a one time tape. To virtually eliminate any possibility of an enemy predicting the chaotic sequence (generated by block 1 in figure 7.5), two recursive chaotic dynamical systems are connected together. One system exhibits chaotic dynamics when iterated in the forward direction, while the other exhibits chaotic dynamics when iterated in the reverse direction. The numbers generated by the two dynamical systems are combined to form the output of block 1 in figure 7.5. The Lyapunov exponents of the dynamical systems, which are chaotic in forward and reverse time respectively, are the reciprocals of each other. This is most easily realized if the equations describing the two dynamical systems are each other's inverse. That is, if one is characterised by $x_{n+1} = f(x_n)$, then the other is characterised by $y_{n+1} = f^{-1}(y_n)$, where $f^{-1}(\cdot)$ denotes the inverse function of $f(\cdot)$.

The numbers generated by a chaotic system have in general a finite length autocorrelation function (Ott 1981, page 656) (i.e. numbers sufficiently far apart in a sequence are effectively independent). While performing computer simulations on chaotic encryption schemes, I found a simple method for transforming sequences of numbers that do not have a flat power spectrum or amplitude pdf into sequences of numbers with a flat power spectrum and amplitude pdf. This transformation consists of three cyclic modulo one adders in series (refer to §7.3 and to §7.5). In addition, this transformation performs all the other functions represented by blocks 2,3 and 4 in figure 7.5.

Sequences generated by an ideal pseudo-random number generator possess maximum entropy. The greater the entropy of the sequence, the less redundancy it exhibits (i.e. the sequence is more random, in the algorithmic sense, than a sequence which possesses less entropy; refer to §3.1). To determine how well the chaotic pseudo-random number generator presented in §7.3 reaches the ideal, two tests are reported (see figure 7.8) on the numbers thereby generated. These tests show that the chaotic pseudo-random number generator performs consistently with an ideal pseudo-random number generator, but cannot of course demonstrate complete equivalence because of the necessarily finite lengths of the sequences.

In influenced chaotic encryption, each bit of the plaintext perturbs each iterate of a chaotic dynamical system. The output from the dynamical system, after undergoing the same transformations as for isolated chaotic encryption, is the cryptogram. The receiver decrypts the cryptogram by performing the inverse transformation to the transmitter. If influenced chaotic encryption is to be useful, the statistics of the cryptogram sequence must be independent of the message sequence, and the cryptogram must possess maximum entropy. To demonstrate how well the influenced chaotic encryption scheme presented in §7.4 achieves this, the same tests as mentioned in the previous paragraph performed on cryptograms generated from a

range of different plaintexts. Plaintexts consisting of repeating sequences of different lengths (repetition lengths from 1 to 16 bits), and plaintexts consisting of finite length sequences (lengths from 1 to 16 bits) having nonequal probability of occurrence were tried. In each case (see figure 7.10) the tests are consistent with statistics of the cryptogram being independent of the plaintext, but do not of course completely confirm this because of the necessarily finite lengths of the sequences.

In reality, a chaotic pseudo-random number generator cannot generate arbitrarily long sequences possessing maximum entropy, since this requires the source to generate an arbitrary large amount of information. A chaotic source can, however, be completely specified by a small amount of information (i.e. the information required to specify the equations and the initial condition is relatively small). Any sequence generated by a chaotic pseudo-random number generator can be specified using this information. The precision of the computation and the amount of information contained in the initial condition determines the amount of information required to specify the generated sequence. In principle, the amount of this information can be made arbitrarily large. The same trade off (i.e. the simplicity of the pseudo-random generator verses the amount of information required to specify the generated sequence) occurs in all encryption schemes, since they are all deterministic. However, the advantage of deterministic chaos over other schemes may be the ease of implementation, and the difficulty of deriving the initial condition (i.e. the difficulty of breaking the encryption) from observation of the cryptogram.

The preliminary studies presented in §7.3 and §7.4 suggest that isolated and influenced chaotic encryption are worthy of a more detailed analysis. Sequences exhibiting maximal entropy are required for encryption. The calculation of entropy for sequences considerably longer than those studied need to be undertaken. This probably requires an analytical approach since simulations are computationally expensive. The chaotic systems ideally suited to encryption are those that generate maximal length sequences when restricted to finite precision numbers. The development of a method which enables one to synthesis such chaotic systems would represent a useful area for further work. Studying how computation using finite precision numbers affects the dynamics of a chaotic system may shed light on how to develop such a method.

9.4 Packet Switching

Packet switching is a quantised communication technique. If a user wishes to transmit a message longer than an allowed limit, the message must first be quantised into packets (*cf.* Kleinrock 1976; Schwartz 1977; Bertsekas and Gallager 1987). The advantage of this is twofold. Firstly, packets enable the network to perform functions which are impossible in other networks (e.g. users with different connection transmission speeds can be connected together). Secondly, packets enable the network to make efficient use of switching and communication resources, since these resources can be shared between all users (i.e. resources are not permanently allocated to individual users). When a user is not transmitting no resources are committed to this user. These resources are free to be used by another user.

Chapter 8 analyses a packet switching flow control algorithm which can become chaotic under certain input traffic patterns, but nevertheless offers significant poten-

tial practical advantages. The algorithm is based on a modification to an algorithm which forms the basis of many flow control algorithms used in practice (Gerla and Kleinrock 1980). The algorithm controls the rate at which acknowledgement packets are sent back to users. The acknowledgement rate is a function of the rate at which packets are received by the network from users. When the packet arrival rate from users into the network is high, the acknowledgement departure rate from the network back to users is low. A user obtains the highest possible instantaneous packet throughput (refer to third paragraph in Chapter 8) by sending one packet immediately after the other into the network. However, the user can only do this for a small number (say from 2 to 5) packets. The network prevents (by delaying the sending of acknowledgements back to the user) the user from sending another packet into the network for a relatively long time. This gives a poor average (i.e. long term) packet throughput. To achieve the highest possible instantaneous throughput, the user pays a penalty in average packet throughput. A user obtains the highest possible average packet throughput by sending packets into the network separated by a suitable time interval. Thus users must make definite decisions as to what strategies best suit their particular needs.

Penalising users for operating an inappropriate packet departure strategy promotes good behaviour by users. For a user to achieve the best performance from the network, a particular packet departure strategy must be adopted. A user operating an inappropriate packet departure strategy is (possibly severely) penalised. Large file transfers, for example, require the network to allocate resources for a long time. The most efficient way to handle a file transfer is for the network to permanently allocate a small portion of its resources to the transfer, and for those resources to be continually in use. This requires the network and the user to form a partnership, each helping the other. The user must provide a regular supply of packets at an appropriate rate while the network must permanently allocate a portion of its resources to the user. To handle bursty traffic (e.g. enquiry-response) efficiently, the network must provide a pool of resources that can be called upon to handle a burst of packets from any user. Some or all of these resources are idle at times and therefore bursty traffic cannot be handled as efficiently as traffic which arrives at regular intervals. The flow control algorithm developed in §8.2 discourages bursty traffic and encourages packets to be sent at regular time intervals.

A packet switching network is a driven deterministic dynamical system. Input packets provide the equivalent of the driving force used in driven dissipative dynamical systems. The bifurcation diagram shown in figure 8.5 is obtained for a sinusoidal input packet arrival pattern. It characterises extraordinarily rich behaviour, between regions of apparent chaos are regions of regularity. The geometrical shape plotted in figure 8.6, formed by constructing a 2-dimensional state space, does not exhibit self similar scaling (i.e. is not fractal). The non-fractal nature may indicate that the geometrical shape cannot be adequately represented in a 2-dimensional state space. The geometric shape thickens as the number of plotted points increases, suggesting that the trajectories do not repeat.

In the remaining paragraphs of this section I indicate what I consider are the most promising areas for future investigation and further development of the techniques introduced and discussed in Chapter 8.

The requirement for ever increasing bandwidth in communication networks re-

mains unabated. The existing design approach is to pack ever more resources into the network, taxing existing technology to the limit. However, an alternative and/or complementary approach is to make use of network resources more efficiently. One way of achieving this is to encourage users to operate a deterministic packet departure process. This enables a network to predict the future arrival time of packets, allowing resources to be allocated and released more efficiently. The flow control algorithm developed in §8.2 was originally conceived as a means of encouraging users (by offering an improved grade of service (GOS), refer to §8.3) to operate a deterministic packet departure process. Users maximise their GOS by sending packets at regular time intervals (a deterministic process). However, under certain packet arrival patterns the flow control algorithm behaves chaotically. Nevertheless, there can be compensatory advantages, as indicated in the following paragraph.

Encouraging certain types of user behaviour represents a means of improving the utilisation of resources within a communication network. However, it also results in the user behaviour being influenced or controlled by the network. Flow control provides a feedback mechanism which may induce chaotic or other forms of apparently self-organising behaviour. Certain forms of such behaviour may maximise the utilisation of network resources. The development of a packet entry flow control algorithm that induces resource efficient self-organising behaviour, while at the same time delivering (close to) the specified GOS to users, may inspire considerable advances in communication network design.

To analyse and take advantage of self-organisation in communication networks requires the development of new design and analysis techniques. Self-organised behaviour is holistic, in the sense that it is the whole activity that is of interest, not the state of any specific subsystem (i.e. it is not possible to reduce the system, or a pattern of activity of the system, to the sum of activities of individual subsystems). It may eventually prove feasible to design a communication network to operate in a specific self-organising coherent or chaotic way to enhance network performance and improve network fault tolerance.

Managing a packet switching network requires accurate knowledge of the user demand, network capacity and network performance. There must therefore be continual measurement and monitoring of the network. The performance of PACNET (a low speed public packet switching network operated by Telecom Corporation of New Zealand Ltd.) is specified by a particular GOS. A network is considered to be well managed if the network capacity is set close to the user demand, while still achieving the specified GOS. The resources required in a packet switching network depends on the user traffic statistics and on the GOS. New Zealand Telecom's ultimate aim is three fold; 1) to accurately measure the performance of PACNET, 2) to determine the statistics of user traffic, and 3) to investigate how traffic statistics affect the amount or resources required for the network. §8.3 describes and demonstrates instruments capable of achieving aims 1) and 2).

Before the statistics of a stream of packets can be measured, it is necessary to develop a model for the packet generation process. A traffic model characterises traffic statistics with a number of random variables. The advent of type-ahead buffering (which allows a user to continue typing characters before the computer has responded to a previous set of characters), intelligent terminals and sophisticated application programs, has made the relationship between the traffic process and the user process

less direct. Little seems to have been written in the archival literature on the development of traffic models for such traffic processes (Pawlita 1981; Jain and Pouther 1986). A suitable traffic model for PACNET users is developed in §8.3.2. Measuring traffic statistics means recording sufficient information that the probability distribution functions of the random variables specified in the traffic model can be estimated.

It is explained in §8.3 how user connections to PACNET were monitored for the purpose of measuring user traffic statistics. An appropriate packet switch traffic measuring instrument (TMI) was developed. Information about the packets that flowed on each of the monitored user connections were recorded for one week. A number of measurements of the packet flow for each user were extracted from the recorded information. The results of these measurements are presented in figure 8.10 - figure 8.16. These preliminary traffic results indicate that a batch Poisson traffic arrival process seems an appropriate model for the four connections examined. Most analytic packet switching models assume a Poisson packet arrival process. However it is well known (*cf.* Chu 1970) that a Poisson approximation to a batch Poisson process seriously underestimates the required buffer and link resources required in a packet switching network. This indicates the importance of measuring traffic statistics.

Analytic models are at present inadequate for predicting the performance of real-world communication networks. Computer simulations will remain a cornerstone for analysing communication networks, at least in the immediate future. Comprehensive simulations investigating how user traffic statistics affect the amount of resources required in a network for a given GOS are lacking. Knowledge of accurate user traffic statistics are necessary to manage a network well. This is particularly important when one remembers that many new services with unknown traffic statistics are continually being connected to packet switching networks. The development of suitable traffic measuring equipment is required if accurate traffic measurements are to be made. The traffic measuring instrument (TMI) described in §8.3.1 should serve as a basis for the design of apparatus for such traffic measurement and for the development of more elaborate instruments incorporating graphical display of real time network performance and traffic statistics.

References

[Abraham and Marsden 1978] R. H. Abraham and J. E. Marsden. *Foundations of Mechanics.* Benjamin: Reading, Massachusetts, 2nd edition, 1978.

[Abraham and Shaw 1987] R. H. Abraham and C. D. Shaw. Dynamics a visual introduction. In F. E. Yates, editor, *Self-Organizing Systems*, pages 543–597, Plenum, 1987.

[Ackerhalt *et al.* 1985] J. R. Ackerhalt, P. W. Milonni, and M. L. Shih. Chaos in quantum optics. *Physics Reports*, 128(4-5):205–300, November 1985.

[Alekseev and Yakobson 1981] V. M. Alekseev and M. V. Yakobson. Symbolic dynamics and hyperbolic dynamic systems. *Physics Reports*, 75(5):287–325, September 1981.

[Allwright 1978] D. J. Allwright. Hypergraphic functions and bifurcations in recurrence relations. *SIAM Journal on Applied Mathematics*, 34(4):687–691, June 1978.

[Ambrozy 1982] A. Ambrozy. *Electrical Noise.* McGraw Hill: New York, 1982.

[Arecchi and Harrison 1987] F. T. Arecchi and R. G. Harrison, editors. *Instabilities and Chaos in Quantum Optics. Springer series in Synergetics 35*, Springer-Verlag: Berlin, 1987.

[Arecchi *et al.* 1982] F. T. Arecchi, R. Meucci, G. Puccioni, and J. Tredicce. Experimental evidence of subharmonic bifurcations, multistability, and turbulence in a Q-switched gas laser. *Physical Review Letters*, 49(17):1217–1220, 25 October 1982.

[Arnold 1963] V. I. Arnold. Small denominators and problems of stability of motion in classical and celestial mechanics. *Russian Mathematical Surveys*, 18(6):85–191, November-December 1963.

[Arnold 1973] V. I. Arnold. *Ordinary Differential Equations.* MIT Press, 1973.

[Arnold 1978] V. I. Arnold. *Mathematical Methods of Classical Mechanics.* Springer-Verlag, 1978.

[Arnold 1983] V. I. Arnold. *Geometrical Methods in the Theory of Differential Equations.* Springer-Verlag, 1983.

[Atkinson 1978] K. E. Atkinson. *An Introduction to Numerical Analysis.* John Wiley & Sons, 1978.

[Auslander *et al.* 1964] J. Auslander, N. P. Bhatia, and P. Seibert. Attractors in dynamical systems. *Biological Soc. Mat. Mex.*, 9:55–66, 1964.

[Babloyantz and Destexhe 1988] A. Babloyantz and A. Destexhe. Is the normal heart a periodic oscillator? *Biological Cybernetics*, 58(3):203–211, 1988.

[Bak 1982] P. Bak. Commensurate phases, incommensurate phases and the devil's staircase. *Reports on the Progress of Physics*, 45(6):587–629, June 1982.

[Bak 1986] P. Bak. The devil's staircase. *Physics Today*, 39(12):38–45, December 1986.

[Barnett and Chen 1986] W. Barnett and P. Chen. Deterministic chaos and fractal attractors as tools for nonparametric dynamical econometric inference. *CER Dissusion Paper 86-6*, 1–33, 13 May 1986.

[Barrow 1982] J. D. Barrow. Chaotic behaviour in general relativity. *Physics Reports*, 85(1):1–49, 1982.

[Bates and McDonnell 1986] R. H. T. Bates and M. J. McDonnell. *Image Restoration and Reconstruction*. Clarendon Press: Oxford, 1986.

[Bates and Murch 1987] R. H. T. Bates and A. R. Murch. Deterministic-chaotic variably coloured noise. *Electronics Letters*, 23(19):995–996, 10 September 1987.

[Beker and Piper 1982] H. Beker and F. Piper. *Cipher Systems: The Protection of Communications*. John Wiley & Sons: New York, 1982.

[Bell 1960] D. A. Bell. *Electrical Noise, Fundamentals and Physical Mechanisms*. Van Nostrard: London, 1960.

[Belyayev *et al.* 1985] R. V. Belyayev, G. M. Vorontsov, N. N. Zalogin, and Y. Kislov. Numerical modeling of stochastic processes in a delayed and amplitude-limited self-excited oscillator. *Soviet Journal of Communications Technology and Electronics*, 30(6):52–59, June 1985.

[Bennet 1956] W. R. Bennet. Methods for solving noise problems. *Proceedings of the IRE*, 44(5):609–638, May 1956.

[Berry *et al.* 1987] M. V. Berry, I. C. Percival, and N. O. Weiss, editors. Special issue on 'dynamical chaos'. *Proceedings of the Royal Society of London, Series A*, 413(1844):1–199, September 1987.

[Bertsekas and Gallager 1987] D. Bertsekas and R. Gallager. *Data Networks*. Prentice-Hall, international edition, 1987.

[Bertsekas *et al.* 1984] D. P. Bertsekas, E. Gafni, and R. G. Gallager. Second derivative algorithms for minimum delay distributed routing in networks. *IEEE Transactions on Communications*, COM-32(8):911–919, August 1984.

[Beurle 1956] R. L. Beurle. A comparison of the noise, and random frequency and amplitude fluctuations in different types of oscillators. *IEE Proceedings*, 103(8, pt. B):182–189, March 1956.

[Binzel *et al.* 1986] R. P. Binzel, J. R. Grenn, and C. B. Opal. Chaotic rotation of Hyperion? *Nature*, 320(6062):511, 10-16 April 1986.

[Birkhoff 1927] G. D. Birkhoff. *Dynamical Systems*. American Mathematical Society, 1927.

[Birkhoff 1941] G. D. Birkhoff. Some unsolved problems of theoretical dynamics. *Science*, 94:598–600, 1941.

[Bowen 1980] R. Bowen. *On Axiom A Diffeomorphisms*. Volume 35 of *Conference Board of the Mathematical Sciences*, American Mathematical Society, 1980.

[Boyer 1968] C. B. Boyer. *A History of Mathematics*. John Wiley & Sons, 1968.

[Bracewell 1978] R. N. Bracewell. *The Fourier Transform and its Applications*. McGraw-Hill, 2nd edition, 1978.

[Braun 1983] M. Braun. *Differential Equations and Their Applications*. Volume 15 of *Applied Mathematical Sciences*, Springer-Verlag, 1983.

[Brickell and Odlyzko 1988] E. F. Brickell and A. M. Odlyzko. Cryptanalysis: A survey of recent results. *Proceedings of the IEEE*, 76(5):578–593, May 1988. Contained in [Simmons 1988].

[Brockett 1982] R. W. Brockett. On conditions leading to chaos in feedback systems. In *Proceedings of the* 21^{th} *IEEE Conference on Decision and Control*, pages 932–936, 1982.

[Butcher *et al.* 1979] J. C. Butcher, K. Burrage, and F. H. Chipman. *Stride: Stable Runge-Kutta integrator for differential equations. Computational Mathematics No. 20*, University of Auckland, Auckland, New Zealand, August 1979.

[Campbell and Rose 1983] D. Campbell and H. Rose, editors. Special issue on 'order in chaos'. *Physica*, 7D(1-3), May 1983.

[Campbell *et al.* 1985] D. Campbell, J. Crutchfield, D. Farmer, and E. Jen. Experimental mathematics: The role of computation in nonlinear science. *Communications of the ACM*, 28(4):374–384, 1985.

[Carr 1981] J. Carr. *Applications of Centre Manifold Theory.* Volume 35 of *Applied Mathematical Sciences*, Springer-Verlag, 1981.

[Cartwright and Littlewood 1945] M. L. Cartwright and J. E. Littlewood. On nonlinear differential equations of the second order. *Journal of the London mathematical Society*, 20:180–189, 1945.

[Casti 1982] J. L. Casti. Recent development and future perspectives in nonlinear system theory. *SIAM Review*, 42(2):301–331, July 1982.

[Chaitin 1975] G. J. Chaitin. Randomness and mathematical proof. *Scientific American*, 232(5):47–52, May 1975.

[Chandra and Scott 1981] J. Chandra and A. C. Scott, editors. *Coupled Nonlinear Oscillators*, Proceedings Joint U.S. Army Center for Nonlinear Studies Workshop, North Holland, 21-23 July 1981.

[Chernikov *et al.* 1988] A. A. Chernikov, R. Z. Sagdeev, and G. M. Zaslavsky. Chaos: How regular can it be? *Physics Today*, 41(11):27–35, November 1988.

[Chi 1966] A. R. Chi, editor. Special issue on 'Frequency stability'. *Proceedings of the IEEE*, 54(2), February 1966.

[Chialvo and Jalife 1987] D. R. Chialvo and J. Jalife. Nonlinear dynamics of cardiac excitation and impulse propagation. *Nature*, 330(6150):749–752, 24-31 December 1987.

[Chillingworth 1976] D. R. J. Chillingworth. *Differential Topology with a view to Applications.* Volume 9 of *Research Notes in Mathematics*, Pitman: London, 1976.

[Chinn and Steenrod 1966] W. G. Chinn and N. E. Steenrod. *First Concepts of Topology.* Random House, 1966.

[Chu 1970] W. W. Chu. Buffer behaviour for batch poisson arrivals and single constant output. *IEEE Transactions on Communications Technology*, COM-18(5):613–618, October 1970.

[Chua and Lin 1988] L. O. Chua and T. Lin. Chaos in digital filters. *IEEE Transactions on Circuits and Systems*, 35(6):648–658, June 1988.

[Chua 1984] L. O. Chua. Nonlinear circuits. *IEEE Transactions on Circuits and Systems*, CAS-31(1):69–87, January 1984.

[Chua 1987] L. O. Chua, editor. Special issue on 'chaotic systems'. *Proceedings of the IEEE*, 75(8), August 1987.

[Chua *et al.* 1983a] L. O. Chua, R. K. Brayton, J. Guckenheimer, and A. I. Mees, editors. Special issue on 'nonlinear phenomena, nonlinear modelling, and nonlinear mathematics'. *IEEE Transactions on Circuits and Systems*, CAS-30(8), August 1983.

[Chua *et al.* 1983b] L. O. Chua, R. K. Brayton, J. Guckenheimer, and A. I. Mees, editors. Special issue on 'nonlinear phenomena, nonlinear modelling, and nonlinear mathematics'. *IEEE Transactions on Circuits and Systems*, CAS-30(9), September 1983.

[Chua *et al.* 1986] L. O. Chua, M. Komuro, and T. Matsumoto. The double scoll family. *IEEE Transactions on Circuits and Systems*, CAS-33(11):1072–1118, November 1986.

[Coates *et al.* 1988] R. F. W. Coates, G. J. Janacek, and K. V. Lever. Monte carlo simulation and random number generation. *IEEE Journal on Selected Areas in Communication*, 6(1):58–66, January 1988.

[Coddington and Levinson 1955] E. Coddington and N. Levinson. *Theory of Ordinary Differential Equations.* McGraw Hill: New York, 1955.

[Collet and Eckmann 1980] P. Collet and J. Eckmann. *Iterations Maps on the Interval as Dynamical Systems. Progress in Physics*, Birkhauser: Boston, 1980.

[Contopoulos 1985] G. Contopoulos. The transition to chaos in galactic models of two and three degrees of freedom. In J. R. Buchler, editor, *Chaos in Astrophysics*, pages 259–271, Reidel, 1985.

[Coveney 1988] P. V. Coveney. The second law of thermodynamics: Entropy irreversibility and dynamics. *Nature*, 333(6172):409–415, 2nd June 1988.

[Crutchfield and Huberman 1980] J. P. Crutchfield and B. A. Huberman. Fluctuations and the onset of chaos. *Physics Letters*, 77A(6):407–410, 23 June 1980.

[Crutchfield *et al.* 1982] J. P. Crutchfield, J. D. Farmer, and B. A. Huberman. Fluctuations and simple chaotic dynamics. *Physics Reports*, 92(2):46–82, December 1982.

[Crutchfield *et al.* 1986] J. P. Crutchfield, J. D. Farmer, N. H. Packard, and R. S. Shaw. Chaos. *Scientific American*, 255(6):38–49, December 1986.

[Cvitanovic 1984] P. Cvitanovic, editor. *Universality in chaos (a reprint selection).* Adam Hilger, 1984. A reprint collection.

[Dauben 1983] J. W. Dauben. Georg Cantor and the origins of transfinite set theory. *Scientific American*, 248(6):112–121, June 1983.

[Davies 1987] P. Davies. The creative cosmos. *New Scientist*, 23(12):41–44, 17 December 1987.

[Davies 1989] P. Davies. *The Cosmic Blueprint.* Unwin Hyman Ltd., First published paperback 1989.

[Devaney 1987] R. L. Devaney. *An Introduction to Chaotic Dynamical Systems.* Addison-Wesley, 1987.

[Diffe and Hellman 1976] W. Diffe and M. E. Hellman. New directions in cryptography. *IEEE Transactions on Information Theory*, IT-22(6):644–654, November 1976.

[Diffe and Hellman 1979] W. Diffe and M. E. Hellman. Privacy and authentication: An introduction to cryptography. *Proceedings of the IEEE*, 67(3):397–427, March 1979.

[Dold and Eckmann 1977] A. Dold and B. Eckmann, editors. *Turbulence Seminar*, Lecture Notes in Mathematics, Springer-Verlag, 1977. Vol 615.

[Dragt and Finn 1976] A. J. Dragt and J. M. Finn. Insolubility of trapped particle motion in a magnetic dipole field. *Journal of Geophysical Research*, 81(13):2327–2340, 1 May 1976.

[Dudick *et al.* 1971] A. L. Dudick, E. Fuchs, and P. E. Jackson. Data traffic measurement for inquiry-response computer communication systems. In C. V. Freiman, editor, *Information Processing 71*, pages 634–641, North-Holland, August 1971.

[Dugas 1957] R. Dugas. *A History of Mechanics.* Routledge & Kegan Paul Ltd, 1957.

[Dutta and Horn 1981] P. Dutta and P. M. Horn. Low-frequency fluctuations in solids: 1/f noise. *Reviews of Modern Physics*, 53(3):497–516, July 1981.

[Eckmann and Ruelle 1985] J. P. Eckmann and D. Ruelle. Ergodic theory of chaos and strange attractors. *Reviews of Modern Physics*, 57(3, pt. 1):617–656, July 1985.

[Eckmann 1981] J. P. Eckmann. Roads to turbulence in dissipative dynamical systems. *Reviews of Modern Physics*, 53(4, pt. 1):643–654, October 1981.

[Endo and Chua 1988] T. Endo and L. O. Chua. Chaos from phase-locked loops. *IEEE Transactions on Circuits and Systems*, CAS-35(8):987–1003, August 1988.

[Esande 1985] D. F. Esande. Stochasticity in classical Hamiltonian systems: Universal aspects. *Physics Reports*, 121(3-4):165–261, May 1985.

[Eves 1969] H. Eves. *An Introduction to the History of Mathematics.* Holt, Rinehart and Winston, 3rd edition, 1969.

[Farmer 1982] J. D. Farmer. Chaotic attractors of an infinite-dimensional dynamical system. *Physica*, 4D(3):366–393, March 1982.

[Farmer *et al.* 1983] J. D. Farmer, E. Ott, and J. A. Yorke. The dimension of chaotic attractors. *Physica*, 7D(1-3):153–180, May 1983. Contained in [Campbell and Rose 1983].

[Farmer *et al.* 1984] D. Farmer, T. Toffoli, and S. Wolfram, editors. Special issue on 'cellular automata'. *Physica*, 10D(1-2), January 1984.

[Feigenbaum 1978] M. J. Feigenbaum. Quantitative universality for a class of nonlinear transformations. *Journal Statistical Physics*, 19(1):25–52, 1978. Reprinted in [Hao 1984].

[Feigenbaum 1979] M. J. Feigenbaum. The universal metric properties of nonlinear transformations. *Journal Statistical Physics*, 21(6):669–706, 1979. Reprinted in [Hao 1984].

[Feigenbaum 1980] M. J. Feigenbaum. Universal behaviour in nonlinear systems. *Los Alamos Science*, 1:4–27, 1980. Reprinted in [Cvitanovic 1984].

[Feigenbaum *et al.* 1982] M. J. Feigenbaum, L. P. Kadanoff, and S. J. Shenker. Quasiperiodicity in dissipative systems: A renormalization group analysis. *Physica*, 5D(2-3):370–386, September 1982.

[Feller 1966] W. Feller. *An Introduction to Probability Theory and its Applications.* Volume 1-2, John Wiley & Sons, 1966.

[Flaschka and Chirikov 1988] H. Flaschka and B. Chirikov, editors. Special issue on 'progress in chaotic dynamics'. *Physica*, 33D(1-3), October-November 1988.

[Ford 1983] J. Ford. How random is a coin toss? *Physics Today*, 36(4):40–47, April 1983.

[Fourier 1822] J. Fourier. *Théorie analytique de la chaleur.* Paris, 1822.

[Franaszek 1984] M. Franaszek. Effect of random noise on the deterministic chaos in a dissipative system. *Physics Letters*, 105A(8):383–386, 5 November 1984.

[Frederickson *et al.* 1983] P. Frederickson, J. L. Kaplan, E. D. Yorke, and J. A. Yorke. The Lyapunov dimension of strange attractors. *Journal of Differential Equations*, 49(2):185–207, August 1983.

[Freire *et al.* 1984] E. Freire, L. G. Franquelo, and J. Araci. Periodicity and chaos in an autonomous electronic system. *IEEE Transactions on Circuits and Systems*, CAS-31(3):237–247, March 1984.

[Galileo 1638] G. Galileo. *Discorsi e Dimostrazioni Mathematiche, intorno á due nuoue scienze.* Translated into English, (Dover: New York, 1914), 1638.

[Gallager 1977] R. G. Gallager. A minimum delay routing algorithm using distributed computation. *IEEE Transactions on Communications*, COM-25(1):73–85, January 1977.

[Gardiner 1983] C. W. Gardiner. *Handbook of Stochastic Methods.* Springer-Verlag, 1983.

[Georganas 1980] N. D. Georganas. Modelling and analysis of message switched computer communication networks with multilevel flow control. *Computer Networks*, 4(6):285–294, December 1980.

[Gerla and Kleinrock 1980] M. Gerla and L. Kleinrock. Flow control: A comparative survey. *IEEE Transactions on Communications*, COM-28(4):553–574, April 1980.

[Giessler *et al.* 1978] A. Giessler, J. Hänle, A. König, and E. Pade. Free buffer allocation - an investigation by simulation. *Computer Networks*, 2(4):191–208, August 1978.

[Gilmore 1981] R. Gilmore. *Catastrophe Theory for Scientists and Engineers.* John Wiley & Sons, 1981.

[Glass *et al.* 1983] L. Glass, M. R. Guevara, A. Shrier, and R. Perez. Bifurcation and chaos in a periodically stimulated cardiac oscillator. *Physica*, 7D(1-3):89–101, May 1983. Contained in [Campbell and Rose 1983].

[Glass *et al.* 1987] L. Glass, A. L. Goldberger, M. C. =, and A. Shrier. Nonlinear dynamics, chaos and complex cardiac arrhythmias. *Proceedings of the Royal Society London Series A*, 413(1844):9–26, September 1987. Contained in [Berry *et al.* 1987].

[Glass *et al.* 1988] L. Glass, A. Beuter, and D. Larocque. Time delay, oscillations, and chaos in physiological control systems. In *Mathematical Biosciences 89:*, Elsevier Science, 1988.

[Glazier and Libchaber 1988] J. A. Glazier and A. Libchaber. Quasi-periodicity and dynamical systems: An experimentalist's view. *IEEE Transactions on Circuits and Systems*, CAS-35(7):790–809, July 1988. Contained in [Kuo 1988].

[Gleick 1987] J. Gleick. *Chaos: Making a New Science.* Viking: New York, 1987.

[Golay 1964] M. J. E. Golay. Normalized equations of the regenerative oscillator – noise, phase-locking, and pulling. *Proceedings of the IEEE*, 52(11):1311–1330, November 1964.

[Goldberger and Rigney 1988] A. L. Goldberger and D. R. Rigney. Sudden death is not chaos. In J. A. S. Kelso, A. J. Mandell, and M. F. Shlesinger, editors, *Dynamic Patterns in Complex Systems*, World Scientific, 1988.

[Goldberger and West 1987] A. L. Goldberger and B. J. West. Chaos in physiology: Health or disease. In H. Degn, A. V. Holden, and L. F. Olsen, editors, *Chaos in Biological Systems*, pages 1–4, Plenum, 1987.

[Goldberger *et al.* 1985] A. L. Goldberger, B. J. West, and V. Bhargava. Nonlinear mechanisms in physiology and pathophysiolgy: Toward a dynamical theory of health and disease. In *Pro. 11th Int. Ass. Math. Com. Sim.*, pages 1–4, World Congress: Oslo, 1985.

[Golubitsky 1978] M. Golubitsky. An introduction to catastrophe theory and its applications. *SIAM Reviews*, 20(2):352–387, April 1978.

[Goodman and Warner 1964] L. E. Goodman and W. H. Warner. *Dynamics.* Blackie & Sons, 1964.

[Graham 1984] R. Graham. Quantum noise and strange attractors. *Physics Reports*, 103(1-4):143–149, 1984. Contained in [Itzykson *et al.* 1984].

[Grassberger and Procaccia 1983] P. Grassberger and I. Procaccia. Measuring the strangeness of strange attractors. *Physica*, 9D(1-2):189–208, October 1983.

[Grassberger and Procaccia 1984] P. Grassberger and I. Procaccia. Dimensions and entropies of strange attractors from a fluctuating dynamics approach. *Physica*, 13D(1-2):34–54, August 1984.

[Grebogi *et al.* 1983] C. Grebogi, E. Ott, and J. A. Yorke. Crises, sudden changes in chaotic attractors, and transient chaos. *Physica*, 7D(1-3):181–200, May 1983. Contained in [Campbell and Rose 1983].

[Grebogi *et al.* 1984] C. Grebogi, E. Ott, S. Pelikan, and J. A. Yorke. Strange attractors that are not chaotic. *Physica*, 13D(1-2):261–268, August 1984.

[Greene 1979] J. M. Greene. A method for determining a stochastic transition. *Journal of Mathematical Physics*, 20(6):1183–1201, 1979.

[Guckenheimer and Holmes 1983] J. Guckenheimer and P. Holmes. *Nonlinear Oscillators, Dynamical Systems and Bifurcations of Vector Fields.* Volume 42 of *Applied Mathematical Sciences*, Springer-Verlag, 1983.

[Guckenheimer 1973] J. Guckenheimer. Review of Thoms book. *Bulletin of the American Mathematical Society*, 79(5):878–890, September 1973.

[Guckenheimer 1979] J. Guckenheimer. Sensitive dependence on initial conditions for one-dimensional maps. *Communications of Mathematical Physics*, 70(2):133–160, 1979.

[Guckenheimer *et al.* 1977] J. Guckenheimer, G. Oster, and A. Ipakchi. The dynamics of density dependent population models. *Journal of Mathematical Biology*, 4(2):101–147, 1977.

[Gupta 1975] M. S. Gupta. Applications of electrical noise. *Proceedings of the IEEE*, 63(7):996–1010, July 1975.

[Gupta 1977] M. S. Gupta, editor. *Electrical Noise: Fundamentals & Sources.* IEEE Press, 1977. A reprint collection.

[Hafner 1966] E. Hafner. The effects of noise in oscillators. *Proceedings of the IEEE*, 54(2):179–198, February 1966. Contained in [Chi 1966].

[Haken 1975] H. Haken. Analogy between higher instabilities in fluids and lasers. *Physics Letters*, 53A(1):77–78, 19 May 1975.

[Haken 1981] H. Haken, editor. *Chaos and Order in Nature. Proceedings of the International Symposium on Synergetics*, Springer-Verlag, April 27- May 2 1981.

[Handel 1980] P. H. Handel. Quantum approach to 1/f noise. *Physics Review A*, 22(2):745–757, August 1980.

[Hao 1984] B. Hao, editor. *Chaos (a reprint selection)*. World Scientific, 1984.

[Harrison and Biswas 1986] R. G. Harrison and D. J. Biswas. Chaos in light. *Nature*, 321(22):394–401, May 1986.

[Harrison 1988] R. G. Harrison. Dynanical instabilities and chaos in lasers. *Contemporary Physics*, 29(4):341–371, July-August 1988.

[Hausdorff 1957] F. Hausdorff. *Set Theory*. Chelsea, 1957.

[Hénon and Heiles 1964] M. Hénon and C. Heiles. The applicability of the third integral of motion: Some numerical experiments. *The Astronomical Journal*, 69(1):73–79, February 1964.

[Hénon 1976] M. Hénon. A two-dimensional mapping with a strange attractor. *Communications in Mathematical Physics*, 50(1):69–77, 1976.

[Hénon 1982] M. Hénon. On the numerical computation of Poincaré maps. *Physica*, 5D(2-3):412–414, September 1982.

[Herzel and Pompe 1987] H. Herzel and B. Pompe. Effects of noise on a nonuniform chaotic map. *Physics Letters*, 122A(2):121–125, 1 June 1987.

[Hirsch and Smale 1974] M. W. Hirsch and S. Smale. *Differential Equations, Dynamical Systems and Linear Algebra*. Academic Press, 1974.

[Hirsch 1984] M. W. Hirsch. The dynamical systems approach to differential equations. *Bulletin (New Series) of the American Mathematical Society*, 11(1):1–64, July 1984.

[Hofstadter 1981] D. R. Hofstadter. Metamagical themas. Strange attractors: mathematical patterns delicately poised between order and chaos. *Scientific American*, 245(5):16–29, November 1981.

[Holmes and Marsden 1982] P. J. Holmes and J. E. Marsden. Horseshoes in perturbations in Hamiltonian systems with two degrees of freedom. *Communications in Mathematical Physics*, 82(4):523–544, 1982.

[Holmes 1987] P. Holmes. Dynamical systems in chaos: Some recent books. *IMA Journal of Applied Mathematics*, 39:91–98, 1987.

[Holt 1978] C. A. Holt. *Electronic Circuits Digital and Analog*. John Wiley & Sons, 1978.

[Hooge 1969] F. N. Hooge. 1/f noise is no surface effect. *Physics Letters*, 29A(3):139–140, 21 April 1969.

[Hooge 1972] F. N. Hooge. Discussion of recent experiments on 1/f noise. *Physica*, 60(1):130–144, July 1972.

[Hooge 1976] F. N. Hooge. 1/f noise. *Physica*, 83B(1):14–23, May 1976.

[Hooge *et al.* 1981] F. N. Hooge, T. G. M. Kleinpenning, and L. K. J. Vandamme. Experimental studies on 1/f noise. *Reports on the Progress in Physics*, 44(5):479–532, May 1981.

[Huberman *et al.* 1980] B. A. Huberman, J. P. Crutchfield, and N. H. Packard. Noise phenomena in Josephson junctions. *Applied Physics Letters*, 37(8):750–752, 15 October 1980.

[Hunter and Kearney 1983] I. W. Hunter and R. E. Kearney. Generation of random sequences with jointly specified probability density and autocorrelation functions. *Biological Cybernetics*, 47(2):141–146, 1983.

[Ikeda 1979] K. Ikeda. Multiple-valued stationary state and its instability of the transmitted light by a ring cavity system. *Optics Communications*, 30(2):257–261, August 1979.

[Itzykson *et al.* 1984] C. Itzykson, Y. Pomeau, and N. Sourlas. Common trends in particle and condensed matter physics. *Physics Reports*, 103(1-4):81–184, 1984. Special section on 'Turbulence, Chaos and Fractals'.

[Jackson and Stubbs 1969] P. E. Jackson and C. D. Stubbs. A study of multiaccess computer communications. In *AFIP Proceedings of the Spring Joint Computer Conference*, pages 491–504, AFIP press, 14-16 May 1969.

[Jaffe 1981] J. M. Jaffe. Bottleneck flow control. *IEEE Transactions on Communications*, COM-29(7):954–962, July 1981.

[Jaffe 1984] A. Jaffe. Ordering the universe: The role of mathematics. *SIAM Review*, 26(4):473–500, October 1984.

[Jain and Pouther 1986] R. Jain and S. A. Pouther. Packet trains - measurement and a new model for computer network traffic. *IEEE Journal Selected Areas Communications*, SAC-4(6):986–995, September 1986.

[Jain 1986] R. Jain. A timeout-based congestion control scheme for window flow-controlled networks. *IEEE Journal on Selected Areas in Communications*, SAC-4(7):1162–1167, October 1986.

[Jefferies 1986] D. J. Jefferies. Bifurcation to chaos in clocked digital systems containing autonomous timing circuits. *Physics Letters*, 115A(3):89–92, 31 March 1986.

[Jespersen *et al.* 1972] J. L. Jespersen, B. E. Blair, and L. E. Gatterer, editors. Special section on 'Time and frequency: Generation, dissemination, applications'. *Proceedings of the IEEE*, 60(5), May 1972.

[Johnson 1971] J. B. Johnson. Electronic noise: The first two decades. *IEEE Spectrum*, 18(2):42–46, February 1971.

[Kao *et al.* 1988] Y. H. Kao, J. C. Huang, and Y. S. Gou. Routes to chaos in the Duffing oscillator with a single potential well. *Physics Letters*, 131A(2):91–97, 8 August 1988.

[Kaplan and Yorke 1979] J. L. Kaplan and J. A. Yorke. Preturbulence: A regime observed in a fluid flow model of Lorenz. *Communications in Mathematical Physics*, 67(2):93–108, 1979.

[Kennedy and Chua 1986] M. P. Kennedy and L. O. Chua. Van der Pol and chaos. *IEEE Transactions on Circuits and Systems*, CAS-33(10):974–980, October 1986.

[Kermani and Kleinrock 1980] P. Kermani and L. Kleinrock. Dynamic flow control in store-and-forward computer networks. *IEEE Transactions on Communications*, COM-28(2):263–271, February 1980.

[Keshner 1982] M. S. Keshner. 1/f noise. *Proceedings of the IEEE*, 70(3):212–218, March 1982.

[Kislov *et al.* 1979] V. Y. Kislov, N. N. Zalogin, and Y. A. Myasin. Study of stochastic self-oscillatory processes in self-excited oscillators with delay. *Radio Engineering and Electronic Physics*, 24(6):92–101, June 1979.

[Kitano *et al.* 1983] M. Kitano, T. Yabuzaki, and T. Ogawa. Chaos and period bifurcations in a simple acoustic system. *Physical Review Letters*, 50(10):713–716, 7 March 1983.

[Kleinrock and Kermani 1980] L. Kleinrock and P. Kermani. Static flow control in store-and-forward computer networks. *IEEE Transactions on Communications*, COM-28(2):271–279, February 1980.

[Kleinrock 1975] L. Kleinrock. *Queuing Systems Volume I: Theory.* Wiley-Interscience, 1975.

[Kleinrock 1976] L. Kleinrock. *Queuing Systems Volume II: Computer Applications.* Willey-Interscience, 1976.

[Kline 1972] M. Kline. *Mathematical Thought from Ancient to Modern Times.* Oxford University Press, 1972.

[Kloeden and Mees 1985] P. E. Kloeden and A. I. Mees. Chaotic phenomena. *Bulletin Mathematical Biology*, 47(6):697–738, 1985.

[Kloeden 1976] P. E. Kloeden. Chaotic difference equations are dense. *Bulletin of the Australian Mathematical Society*, 15(3):371–379, 1976.

[Kobayashi and Konheim 1977] H. Kobayashi and A. G. Konheim. Queuing models for computer communications systems analysis. *IEEE Transactions on Communications*, COM-25(1):2–29, January 1977.

[Kousik *et al.* 1985] G. S. Kousik, C. M. V. Vliet, G. Bosman, and P. H. Handel. Quantum 1/f noise associated with ionised impurity scattering and electron-phonon scattering in condensed matter. *Advances in Physics*, 34(6):663–702, November-December 1985.

[Kumar 1980] K. B. Kumar. Optimum end-to-end flow control in networks. In *International Communications Conference (ICC) 1980 Proceedings*, Seattle, Washington, June 1980.

[Kuo 1988] Y. L. Kuo, editor. Special issue on 'Chaos and bifurcations of circuits and systems'. *IEEE Transactions on Circuit and Systems*, CAS-35(7), July 1988.

[Labetoulle and Pujolle 1981] J. Labetoulle and G. Pujolle. A study of flows through virtual circuits. *Computer Networks*, 5(2):119–126, April 1981.

[Lam and Wong 1982] S. S. Lam and J. W. Wong. Queuing network models of packet switching networks part 2: Networks with population size constraints. *Performance Evaluation*, 2(3):161–180, 1982.

[Lam 1976] S. S. Lam. Store-and-forward buffer requirements in a packet switching network. *IEEE Transactions on Communications*, COM-24(4):394–403, April 1976.

[Landau and Lifshitz 1959] L. D. Landau and E. M. Lifshitz. *Fluid Mechanics.* Pergamon Press, 1959. Laudau's original paper is reprinted in [Hao 1984].

[La Salle 1976] J. P. La Salle. *The Stability of Dynamical Systems.* Volume 25 of *Reginal Conference Series in Applied Mathematics*, Published by SIAM, 1976.

[Lebowitz and Penrose 1973] J. L. Lebowitz and O. Penrose. Modern ergodic theory. *Physics Today*, 26(2):23–29, February 1973.

[Lefschetz 1950] S. Lefschetz, editor. *Contributions to the Theory of Nonlinear Oscillators.* Princeton University Press: Princeton, 1950.

[Lerche and Low 1982] I. Lerche and B. C. Low. Some nonlinear problems in astrophysics. *Physica*, 4D(3):293–318, March 1982.

[Li and Yorke 1975] T. Y. Li and J. A. Yorke. Period three implies chaos. *The American Mathematical Monthly*, 82(10):985–992, December 1975.

[Lichtenberg and Lieberman 1983] A. J. Lichtenberg and M. A. Lieberman. *Regular and Stochastic Motion.* Volume 38 of *Applied Mathematical Sciences*, Springer-Verlag, 1983.

[Liu 1980] R. Liu, editor. Special section on 'nonlinear circuits and systems'. *IEEE transactions on circuits and systems*, CAS 27(11), November 1980.

[Longe 1983] G. Longe, editor. *Secure Digital Communications. International Centre for Mechanical Sciences*, Springer-Verlag, 1983.

[Lorenz 1963] E. N. Lorenz. Deterministic nonperiodic flow. *Journal Atmospheric Science*, 20, 1963. Reprinted in [Cvitanovic 1984].

[MacDonald 1962] D. K. C. MacDonald. *Noise and Fluctuations: An Introduction.* John Wiley & Sons, 1962.

[MacKay and Glass 1977] M. C. MacKay and L. Glass. Oscillation and chaos in physiological control systems. *Science*, 197(4300):287–289, 15 July 1977.

[MacKay *et al.* 1984] R. S. MacKay, J. D. Meiss, and I. C. Percival. Transport in Hamiltonian systems. *Physica*, 13D(1-2):55–81, August 1984.

[Mandelbrot 1977] B. B. Mandelbrot. *Fractals: Form, change, and dimension.* W. H. Freeman, 1977.

[Mandelbrot 1987] B. B. Mandelbrot. Towards a second stage of indeterminism in science. *Interdisciplinary Science Review*, 12(2):117–127, 1987.

[Manneville and Pomeau 1980] P. Manneville and Y. Pomeau. Different ways to turbulence in dissipative dynamical systems. *Physica*, 1D(2):219–226, June 1980.

[Marion 1970] J. B. Marion. *Classical Dynamics of Particles and Systems.* Academic Press, 2nd edition, 1970.

[Marotto 1978] F. R. Marotto. Snap-back repellers imply chaos in $\mathbf{R}^n$. *Journal of Mathematical Analysis and Applications*, 63(1):199–223, 15 March 1978.

[Marsden and McCracken 1976] J. E. Marsden and M. McCracken. *The Hopf Bifurcation and Its Applications.* Volume 19 of *Applied Mathematical Science*, Springer-Verlag: New York, 1976.

[Mason *et al.* 1986] J. Mason, P. Mathias, and J. H. Westcott, editors. Special issue on 'predicatability in science and society'. *Proceedings of the Royal Society of London, Series A*, 407(1832):1–145, September 1986.

[Massey 1988] J. L. Massey. An introduction to contemporary cryptology. *Proceedings of the IEEE*, 76(5):533–549, May 1988. Contained in [Simmons 1988].

[Matsumoto 1984] T. Matsumoto. A chaotic attractor from Chua's circuit. *IEEE Transactions on Circuit and Systems*, CAS-31(12):1055–1058, December 1984.

[Matsumoto 1987] T. Matsumoto. Chaos in electronic circuits. *Proceedings of the IEEE*, 75(8):1033–1057, August 1987. Contained in [Chua 1987].

[Matsumoto *et al.* 1985] T. Matsumoto, L. O. Chua, and M. Komuro. The double scroll. *IEEE Transactions on Circuits and Systems*, CAS-32(8):797–818, August 1985.

[Matsumoto *et al.* 1986a] T. Matsumoto, L. O. Chua, and K. Kobayashi. Hyperchaos: Laboratory experiment and numerical confirmation. *IEEE Transactions Circuits and Systems*, CAS-33(11):1143–1147, November 1986.

[Matsumoto *et al.* 1986b] T. Matsumoto, L. O. Chua, and M. Komuro. The double scroll bifurcations. *International Journal on Circuit Theory and Applications*, 14(2):117–146, April 1986.

[May 1974] R. M. May. Biological populations with nonoverlapping generations: Stable points, stable cycles, and chaos. *Science*, 186(4164):645–647, 15 November 1974.

[May 1976] R. M. May. Simple mathematical models with very complicated dynamics. *Nature*, 261(5560):459–467, 10 June 1976. Reprinted in [Cvitanovic 1984].

[May 1980] R. M. May. Models for single populations. In R. M. May, editor, *Theoretical Ecology, Principles and Applications*, chapter 2, pages 5–29, Blackwell Scientific, 1980.

[McCauley 1988] J. L. McCauley. *An Introduction to Nonlinear Dynamics and Chaos Theory.* Volume T20, Physica Scripta, 1988.

[McCulloch 1986] S. McCulloch. *IBM PC Interface to the X25 Packet Switching Network.* Technical Report, University of Canterbury, 1986. 3rd Professional Year Project Report.

[McGonigal and Elmasry 1987] G. C. McGonigal and M. I. Elmasry. Generation of noise by electronic iteration of the logical map. *IEEE Transactions on Circuits and Systems*, CAS-34(8):981–983, August 1987.

[Mees and Chua 1979] A. I. Mees and L. O. Chua. The Hopf bifurcation theorem and its applications to nonlinear oscillations in circuit and systems. *IEEE Transactions on Circuit and Systems*, CAS-26(4):235–254, April 1979.

[Mees and Sparrow 1981] A. I. Mees and C. T. Sparrow. Chaos. *IEE Proceedings*, 128(5, pt. D):201–205, September 1981.

[Mees 1981] A. I. Mees. *Dynamics of Feedback Systems.* John Wiley & Sons, 1981.

[Mees 1983] A. I. Mees. A plain man's guide to bifurcations. *IEEE Transactions on Circuit and Systems*, CAS-30(8):512–517, August 1983. Contained in [Chua *et al.* 1983a].

[Meyer and Matyas 1982] C. H. Meyer and S. M. Matyas. *Cryptography: A New Dimension in Computer Data Security.* John Wiley & Sons, 1982.

[Milnor 1985] J. Milnor. On the concept of attractor. *Communications in Mathematical Physics*, 99(2):177–195, 1985.

[Mira 1987] C. Mira. *Chaotic Dynamics From One-Dimensional Endomorphism to the Two-Dimensional Diffeomorphism.* World Scientific, 1987.

[Moll 1955] J. L. Moll. Junction transistor electronics. *Proceedings of the IRE*, 43(12):1807–1819, December 1955.

[Moon 1987] F. C. Moon. *Chaotic Vibrations, An Introduction for Applied Scientists and Engineers.* John Wiley & Sons, 1987.

[Moser 1973] J. Moser. *Stable and Random Motions in Dynamical Systems, With Special Emphasis on Celestial Mechanics. Annals of Mathematics Studies No. 77*, Princeton University Press: New York, 1973.

[Moser 1986] J. Moser. Recent developments in the theory of Hamiltonian systems. *SIAM Reviews*, 28(4):459–485, December 1986.

[Muralidhar and Sundareshan 1986] K. H. Muralidhar and M. K. Sundareshan. On the decomposition of large communication networks for hierarchical control implementation. *IEEE Transactions on Communications*, COM-34(10):985–987, October 1986.

[Muralidhar 1984] K. H. Muralidhar. Adaptive routing and flow control in large communication networks: A hierarchical scheme for multiobjective optimization. In *Proceedings of the IEEE Infocom*, pages 299–308, San Francisco, April 1984.

[Murch and Bates 1987] A. R. Murch and R. H. T. Bates. Non-random noise mechanisms. *Proceedings of the 24th National Electronics Conference*, 24:137–140, 1-3 September 1987. Held at the University of Auckland, Auckland, New Zealand.

[Murch and Bates 1989] A. R. Murch and R. H. T. Bates. Colored noise generation through deterministic chaos. *Accepted for publication in IEEE Transactions on Communications*, 1989.

[Murch *et al.* 1987] A. R. Murch, W. K. Kennedy, and R. Davidson. Traffic and performance measurements on PACNET. *Proceedings of the 24th National Electronics Conference*, 24:107–110, 1-3 September 1987. Held at the University of Auckland, Auckland, New Zealand.

[Nelson 1973] C. R. Nelson. *Applied Time Series Analysis for Managerial Forecasting.* Holden-Day: San Fanciso, 1973.

[Newell 1977] A. C. Newell. Finite amplitude instabilities of partial difference equations. *SIAM Journal Applied Mathematics*, 33(1):133–160, July 1977.

[Newton 1726] I. Newton. *Philosphiae Naturalis Principle Mathematica.* London, 3rd edition, 1726. The Third Edition with Variant Readings, assembled and edited by A Koyré and I. B. Cohen, Cambridge: Massachusetts, 1972.

[Nicolis 1986a] G. Nicolis. Dynamical systems. *Reports on the Progress in Physics*, 49(8):873–949, August 1986.

[Nicolis 1986b] J. S. Nicolis. Chaotic dynamics applied to information processing. *Reports on the Progress in Physics*, 49(10):1109–1196, October 1986.

[Normand *et al.* 1977] C. Normand, Y. Pomeau, and M. G. Velarde. Convective instability: A physicists approach. *Reviews of Modern Physics*, 49(3):581–624, July 1977.

[Osborne *et al.* 1986] A. R. Osborne, A. D. Kirwan, A. Provenzale, and L. Bergamasco. A search for chaotic behaviour in large and mesoscale motions in the Pacific Ocean. *Physica*, 23D(1-3):75–83, December 1986.

[Ott 1981] E. Ott. Strange attractors and chaotic motions of dynamical systems. *Reviews of Modern Physics*, 53(4 pt. 1):655–671, October 1981.

[Ottino *et al.* 1988] J. M. Ottino, C. W. Leong, H. Rising, and P. D. Swanson. Morphological structures produced by mixing in chaotic flows. *Nature*, 333(6172):419–425, 2nd June 1988.

[Pacault and Vidal 1978] A. Pacault and C. Vidal, editors. *Synergetics Far from Equilibrium. Proceedings of the Conference Far from Equilibrium: Instabilities and Structures*, Springer-Verlag, 27-29 September 1978.

[Packard *et al.* 1980] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw. Geometry from a time series. *Physical Review Letters*, 45A(9):712–716, 1 September 1980.

[Papoulis 1965] A. Papoulis. *Probability, Random Variables and Stochastic Processes.* McGraw-Hill, 1965.

[Pars 1962] L. A. Pars. *An Introduction to the Calculus of Variations.* Wiley: New York, 1962.

[Pawlita 1981] P. F. Pawlita. Traffic measurements in data networks, recent measurement results and some implications. *IEEE Transactions on Communications*, COM-29(4):525–535, April 1981.

[Pederson *et al.* 1973] N. F. Pederson, M. R. Samuelsen, and K. Saermark. Parametric excitation of plasma oscillators in Josephson junctions. *Journal Applied Physics*, 44(11):3113–3117, November 1973.

[Peebles 1980] P. Z. Peebles. *Probability, Random Variables, and Random Signal Principles.* McGraw-Hill, 1980.

[Pei *et al.* 1986] L. Q. Pei, F. Guo, S. X. Wu, and L. O. Chua. Experimental confirmation of the period-adding route to chaos in a nonlinear circuit. *IEEE Transactions on Circuits and Systems*, CAS-33(4):438–442, April 1986.

[Peitgen and Richter 1986] H. O. Peitgen and P. H. Richter, editors. *The Beauty of Fractals.* Springer-Verlag, 1986.

[Peixoto 1962] M. M. Peixoto. Structural stability on two-dimensional manifolds. *Topology*, 1:101–120, April-June 1962.

[Penrose 1979] O. Penrose. Foundations of statistical mechanics. *Reports on the Progress in Physics*, 42(12):1937–2006, December 1979.

[Percival and Richards 1982] I. Percival and D. Richards. *Introduction to Dynamics.* Cambridge University Press, 1982.

[Pesin 1977] Y. B. Pesin. Characteristic Lyapunov exponents and smooth ergodic theory. *Russian Mathematical Surveys*, 32(4):55–114, July-August 1977.

[Pierce 1980] J. R. Pierce. *An introduction to information theory.* Dover: New York, 1980.

[Pippard 1985] A. B. Pippard. *Response and Stability an Introduction to the Physical Theory.* Cambridge University Press, 1985.

[Poincaré 1899] H. Poincaré. *Les méthodes nouvelles de la mécanique celeste.* Gauthier-Villars: Paris, (Vol 1) 1892, (Vol 2) 1893, (Vol 3) 1899. Translated in English (Dover: New York, NASA, 1957).

[Pomeau and Manneville 1980] Y. Pomeau and P. Manneville. Intermittent transition to turbulence in dissipative dynamical systems. *Communications on Mathematical Physics*, 74(2):189–197, 1980.

[Popper and Eccles 1977] K. Popper and J. Eccles. *The Self and Its Brain.* Springer International, 1977.

[Poston and Stewart 1978] T. Poston and I. N. Stewart. *Catastrophe theory and its applications.* Pitman, 1978.

[Press 1978] W. H. Press. Flicker noises in astronomy and elsewhere. *Comments on Astrophysics*, 7(4):103–119, 1978.

[Preston 1983] Lecture Notes in Mathematics. *Iterates of Maps on an Interval*, Springer-Verlag: Berlin, 1983. Vol 999.

[Prigogine and Stengers 1984] I. Prigogine and I. Stengers. *Order Out of Chaos.* Heinemann: London, 1984.

[Prigogine 1980] I. Prigogine. *From Being to Becoming: Time and Complexity in the Physical Sciences.* Freeman, 1980.

[Procaccia and Schuster 1983] I. Procaccia and H. Schuster. Functional renormalization group theory of universal 1/f noise in dynamical systems. *Physical Review A*, 28(2):1210–1212, August 1983.

[Procaccia 1988] I. Procaccia. Universal properties of dynamical complex systems: The organisation of chaos. *Nature*, 333(6174):618–623, 16 June 1988.

[Prufer 1985] M. Prufer. Turbulence in multistep methods for initial value problems. *SIAM Journal Applied Mathematics*, 45(1):32–69, February 1985.

[Reid 1975] W. T. Reid. Anatomy of the ordinary differential equation. *The American Mathematical Monthy*, 82(10):971–984, December 1975.

[Reiser 1979] M. Reiser. A queuing network analysis of computer communication networks with window flow control. *IEEE Transactions on Communications*, COM-27(8):1199–1209, August 1979.

[Reiser 1982] M. Reiser. Performance evaluation of data communication systems. *Proceedings of the IEEE*, 70(2):171–196, February 1982.

[Robins 1984] W. P. Robins. *Phase Noise in Signal Sources (Theory and Applications). IEE Telecommunications Series 9*, Peter Peregrinus, 1984. Paperback edition.

[Rodriguez-Vazquez *et al.* 1985] A. B. Rodriguez-Vazquez, J. L. Huertas, and L. O. Chua. Chaos in a switched-capacitor circuit. *IEEE Transactions on Circuits Systems*, CAS-32(10):1083–1085, October 1985.

[Rössler 1976] O. E. Rössler. An equation for continuous chaos. *Physics Letters*, 57A(5):397–398, 12 July 1976.

[Rössler 1979] O. E. Rössler. An equation for hyperchaos. *Physics Letters*, 71A(2-3):155–157, 30 April 1979.

[Rudin and Mueller 1980] H. Rudin and H. Mueller. Dynamic routing and flow control. *IEEE Transactions on Communications*, COM-28(7):1030–1039, July 1980.

[Ruelle and Takens 1971] D. Ruelle and F. Takens. On the nature of turbulence. *Communications in Mathematical Physics*, 20(3):167–192, 1971.

[Ruelle 1980] D. Ruelle. Strange attractors. *The Mathematical Intelligencer*, 2, 1980. Reprinted in [Cvitanovic 1984].

[Ruelle 1981] D. Ruelle. Small random perturbations of dynamical systems and the definition of attractors. *Communications on Mathematical Physics*, 82(1):137–151, 1981.

[Russel *et al.* 1980] D. A. Russel, J. D. Hanson, and E. Ott. Dimension of strange attractors. *Physical Review Letters*, 45(14):1175–1178, 6 October 1980.

[Rutman 1978] J. Rutman. Characterisation of phase and frequency instabilities in precision frequency sources: Fifteen years of progress. *Proceeding of the IEEE*, 66(9), September 1978.

[Saito 1985] T. Saito. A chaos generator based on a quasi-harmonic oscillator. *IEEE Transactions on Circuits and Systems*, CAS-32(4):320–331, April 1985.

[Salam and Sastry 1985] F. M. A. Salam and S. S. Sastry. Dynamics of the forced Josephson junction circuit: The regions of chaos. *IEEE Transactions on Circuits and Systems*, CAS-32(8):784–796, August 1985.

[Sanz-Serma 1987] J. M. Sanz-Serma. Studies in numerical nonlinear instability III: Augmented Hamiltonian systems. *SIAM Journal Applied Mathematics*, 47(1):92–108, February 1987.

[Sauer and Chandy 1981] C. H. Sauer and K. M. Chandy. *Computer Systems Performance Modeling. Prentice-Hall Series in Advances in Computing Science and Technology*, Prentice-Hall: New Jersy, 1981.

[Schuster 1983] H. G. Schuster. *Chaotic Behaviour in Systems.* Physik-Verlag, 1983.

[Schwartz and Stern 1980] M. Schwartz and T. E. Stern. Routing techniques used in computer communication networks. *IEEE Transactions on Communications*, COM-28(4):539–552, April 1980.

[Schwartz 1977] M. Schwartz. *Computer Communication Network Design and Analysis.* Prentice Hall: New Jersy, 1977.

[Shannon 1948] C. E. Shannon. A mathematical theory of communications. *Bell System Technical Journal*, 27(3):379–423, July 1948.

[Shannon 1949] C. E. Shannon. Communication theory of secrecy systems. *Bell System Technical Journal*, 28(4):656–715, October 1949.

[Shaw 1981] R. Shaw. Modelling chaotic systems. In H. Haken, editor, *Chaos and Order in Nature*, chapter 10, pages 218–231, Springer-Verlag, 27 April-2 May 1981.

[Silvestor 1985] M. Silvestor. *Computer Communications Packet Switching Network Tester.* Technical Report, University of Canterbury, 1985. 3rd Professional Year Project Report.

[Simmons 1985] G. J. Simmons. Cryptology. In *Encyclopaedia Britannica*, pages 913–924, Encyclopaedia Britannica: Chicago, 1985.

[Simmons 1988] G. J. Simmons, editor. Special section on 'Cryptology'. *Proceedings of the IEEE*, 76(5), May 1988.

[Singer 1978] D. Singer. Stable orbits and bifurcations of maps of the interval. *SIAM Journal of Applied Mathematics*, 35(2):260–267, September 1978.

[Skilling 1974] H. H. Skilling. *Electrical Networks.* John Wiley & Sons, 1974.

[Smale 1966] S. Smale. Structurally stable systems are not dense. *American Journal of Mathematics*, 88(?):491–495, ? 1966.

[Smale 1967] S. Smale. Differentiable dynamical systems. *Bulletin of the American Mathematical Society*, 73(6):747–817, November 1967.

[Smith 1987] V. A. Smith. *Non Random Noise in Nonlinear Circuits (Deterministic Chaos).* Technical Report, University of Canterbury, 1987. 3rd Professional Year Project Report.

[Sparrow 1980] C. T. Sparrow. Bifurcation and chaotic behaviour in simple feedback systems. *Journal of Theoretical Biology*, 83(1):93–105, 7 March 1980.

[Sparrow 1982] C. Sparrow. *The Lorenz Equations: Bifurcation, Chaos and Strange Attractors.* Springer-Verlay, 1982.

[Sporns *et al.* 1987] O. Sporns, S. Roth, and F. F. Seelig. Chaotic dynamics of two coupled biochemical oscillators. *Physica*, 26D(1-3):215–224, May-June 1987.

[Sproule and Mellor 1981] D. E. Sproule and F. Mellor. Routing, flow, and congestion control in the Datapac network. *IEEE Transactions on Communications*, COM-29(4):386–391, April 1981.

[Stewart 1981] I. Stewart. Applications of catastrophe theory to the physical sciences. *Physica*, 2D(2):245–305, April 1981.

[Stewart 1982] I. Stewart. Catastrophe theory in physics. *Reports on the Progress in Physics*, 45(2):185–221, February 1982.

[Swan 1985] C. Swan. *Packet Switching Test Set.* Technical Report, University of Canterbury, 1985. 3rd Professional Year Project Report.

[Swinney and Gollub 1978] H. L. Swinney and J. P. Gollub. The transition to turbulence. *Physics Today*, 31(8):41–49, August 1978.

[Taub 1985] H. Taub. *Digital Circuits and Microprocessors.* McGraw Hill, international student edition, 1985.

[Teman 1988] R. Teman. *Infinite Dimensional Dynamical Systems in Mechanics and Physics.* Volume 68 of *Applied Mathematical Sciences*, Springer-Verlag: New York, 1988.

[Terman 1982] F. E. Terman. *Electronic and Radio Engineering.* McGraw Hill, 4th edition, 1982.

[Tomita 1982] K. Tomita. Chaotic response of nonlinear oscillators. *Physics Reports*, 86(3):113–167, June 1982.

[Tsitsiklis and Bertsekas 1986] J. N. Tsitsiklis and D. P. Bertsekas. Distributed asynchronous optimal routing in data networks. *IEEE Transactions on Automatic Control*, AC-31(4):325–332, April 1986.

[Tymes 1981] L. W. Tymes. Routing and flow control in TYMNET. *IEEE Transactions on Communications*, COM-29(4):392–398, April 1981.

[Ushio and Hirai 1985] T. Ushio and K. Hirai. Chaotic behaviour in piecewise-linear sampled-data control systems. *International Journal on Nonlinear Mechanics*, 20(5-6):493–506, 1985.

[Ushio and Hsu 1987] T. Ushio and C. S. Hsu. Chaotic rounding error in digital control systems. *IEEE Transactions on Circuits and Systems*, CAS-34(2):133–139, February 1987.

[van der Pol and van der Mark 1927] B. van der Pol and J. N. van der Mark. Frequency demultiplication. *Nature*, 120(3019):363–364, 10 September 1927.

[van der Pol 1934] B. van der Pol. The nonlinear theory of electronic oscillators. *Proceedings of the IRE*, 22(9):1051, September 1934.

[Van der ziel 1988] A. Van der ziel. Unified presentation of 1/f noise in electronic devices: Fundamental 1/f noise. *Proceedings of the IEEE*, 76(3):233–258, March 1988.

[van Kampen 1983] N. G. van Kampen. *Stochastic Processes in Physics and Chemistry.* North-Holland, 1983.

[Van Vliet 1987] C. M. Van Vliet, editor. *Ninth International Conference on Noise in Physical Systems*, Natural Science and Engineering Research Council Ottawa, Canada, World Scientific, 1987.

[Voss and Clarke 1976] R. F. Voss and J. Clarke. Flicker (1/f) noise: Equilibrium temperature and resistance fluctuations. *Physical Review B*, 13(2):556–573, 15 January 1976.

[Walker and Ford 1969] G. H. Walker and J. Ford. Amplitude instability and ergodic behavior for conservative nonlinear oscillator systems. *The Physical Review*, 188(1):416–432, December 1969.

[Webb and Gershenfield 1987] W. W. Webb and N. A. Gershenfield. The dimension of 1/f noise. *Bulletin American Physical Society*, 32(3):482, 1987.

[Weissman 1988] M. B. Weissman. 1/f noise and other slow, nonexponential kinetics in condensed matter. *Reviews of Modern Physics*, 60(2):537–571, April 1988.

[West and Goldberger 1987] B. J. West and A. L. Goldberger. Physiology in fractal dimensions. *American Scientist*, 75(4):354–365, July-August 1987.

[Wiggins 1988] S. Wiggins. *Global Bifurcation and Chaos: Analytical Methods.* Volume 73 of *Applied Mathematical Sciences*, Springer-Verlag, 1988.

[Winfree 1974] A. T. Winfree. Rotating chemical reactions. *Scientific American*, 230(6):82–95, June 1974.

[Wisdom 1982] J. Wisdom. The origin of the Kirkwood gaps: A mapping for asteroidal motion near the 3/1 commensurability. *The Astronomical Journal*, 87(3):577–593, March 1982.

[Wisdom 1987a] J. Wisdom. Chaotic behaviour in the solar system. *Proceedings of the Royal Society of London, Series A*, 412(1844):109–129, 8 September 1987. Contained in [Berry *et al.* 1987].

[Wisdom 1987b] J. Wisdom. Urey price lecture: Chaotic dynamics in the solar system. *ICARUS*, 72(2):241–275, November 1987.

[Wolfram 1984] S. Wolfram. Cellular automata as models of complexity. *Nature*, 311(5985):419–424, 4 October 1984.

[Wolfram 1985a] S. Wolfram. Cryptograph with cellular automata. In H. C. Williams, editor, *Advances in Cryptology - Crypto 84*, pages 429–432, Lecture Notes in Computer Science Vol 218, Springer-Verlag, 18-22 August 1985.

[Wolfram 1985b] S. Wolfram. Origins of randomness in physical systems. *Physics Review Letters*, 55(5):449–452, 29 July 1985.

[Wong 1984] S. Wong. Newtons method and symbolic dynamics. *Proceedings of the American Mathematical Society*, 91(2):245–253, June 1984.

[Woodcock and Davis 1978] A. Woodcock and M. Davis. *Catastrophe Theory.* E. P. Dutton: New York, 1978.

[Wu 1987] S. Wu. Chua's circuit family. *Proceeding of the IEEE*, 75(8):1022–1032, August 1987. contained in [Chua 1987].

[Yamaguti and Ushiki 1981] M. Yamaguti and S. Ushiki. Chaos in numerical analysis of ordinary differential equations. *Physica*, 3D(3):618–626, August 1981.

[Yuen and Fraser 1979] C. K. Yuen and D. Fraser. *Digital Spectral Analysis.* Pitman, 1979.

[Zahler and Sussmann 1977] R. S. Zahler and H. J. Sussmann. Claims and accomplishments of applied catastrophe theory. *Nature*, 269(5631):759–763, 27 October 1977. Correspondence: 1 December, 29 December.

[Zeeman 1976] E. C. Zeeman. Catastrophe theory. *Scientific American*, 234(4), April 1976.

[Zeeman 1977] E. C. Zeeman. *Catastrophe Theory, Selected Papers 1972-1977.* Reading: Benjamin, 1977.

[Zeeman 1983] E. C. Zeeman. *Catastrophe Theory and Applications.* Summer School on Dynamical Systems, August 1983.

[Zhong and Ayrom 1985] G. Q. Zhong and F. Ayrom. Experimental confirmation of chaos from Chua's circuit. *International Journal on Circuit Theory and Applications*, 13(1):93–98, January 1985.